

PTO 04-272

**BEST AVAILABLE COPY**

Japanese Kokai Patent Application  
No. Hei 11[1999]-95793

**SPEECH INPUT INTERPRETER AND SPEECH INPUT INTERPRETATION METHOD**

Takehide Okuno, et al.

UNITED STATES PATENT AND TRADEMARK OFFICE  
WASHINGTON, D.C. NOVEMBER 2003  
TRANSLATED BY THE RALPH MCELROY TRANSLATION COMPANY

JAPANESE PATENT OFFICE  
PATENT JOURNAL (A)  
KOKAI PATENT APPLICATION NO. HEI 11[1999]-95793

|                         |                             |
|-------------------------|-----------------------------|
| Int. Cl. <sup>6</sup> : | G 10 L 3/00<br>5/04<br>9/00 |
| Filing No.:             | Hei 9[1997]-252446          |
| Filing Date:            | September 17, 1997          |
| Publication Date:       | April 9, 1999               |
| No. of Claims:          | 12 (Total of 19 pages; OL)  |
| Examination Request:    | Not requested               |

SPEECH INPUT INTERPRETER AND SPEECH INPUT INTERPRETATION METHOD

[Onsei nyuryoku kaishaku sochi oyobi onsei nyuryoku kaishaku hoho]

|            |                            |
|------------|----------------------------|
| Inventors: | Takehide Okuno, et al.     |
| Applicant: | 000003078<br>Toshiba Corp. |

[There are no amendments to this patent.]

\* \* \*

Claims

1. A speech input interpreter characterized in that in a speech input interpreter which interprets input speech and outputs information on applicable vocabulary items, it is equipped with  
a means which stores first information pertaining to normal vocabulary items and second information which consids the possibility that some of said normal vocabulary items may be replaced by prescribed alternative expressions during speech input,

/2\*

---

\* [Numbers in the margin indicate pagination in the foreign document.]

a means for speech recognition of the input speech,  
a means which detects the aforementioned alternative expression from the result of the aforementioned speech recognition based on the aforementioned second information, and  
a means which runs a query in the aforementioned first information based at least on the vocabulary items other than said alternative expression included in the aforementioned input speech recognition result when the aforementioned alternative expression is detected from the aforementioned recognition result by said means in order to find said applicable vocabulary item.

2. The speech input interpreter described under Claim 1 characterized in that it is further equipped with a means which evaluates the priorities of applicable vocabulary items based at least on the phonemic characteristic of the speech corresponding to the aforementioned alternative expression when multiple applicable vocabulary items are retrieved.

3. A speech input interpreter characterized in that in a speech input interpreter which interprets input speech and outputs information on applicable vocabulary items, it is equipped with

a vocabulary storage means which stores alternative expressions obtained by replacing some prescribed normal vocabulary items as the target of the speech recognition by alternative expressions used to replace arbitrary words as one type of vocabulary item,

a metrical information storage means which stores the notations of and metrical information on the aforementioned normal vocabulary items which do not contain the aforementioned alternative expressions out of the vocabulary items stored in the aforementioned vocabulary storage means,  
a speech analysis means which carries out speech recognition and metrical analysis of speech input through a speech input device in reference to the aforementioned vocabulary storage means, and

a substitution expression verification means which replaces the aforementioned replaced part by the aforementioned normal vocabulary item based on the result of the aforementioned speech recognition of the aforementioned input speech and the result of the aforementioned metrical analysis by the aforementioned speech analysis means in reference to the aforementioned metrical information storage means.

4. A speech input interpreter characterized in that in a speech input interpreter equipped with a means which analyzes speech input from a speech input device for speech recognition and outputs a speech analysis result including a speech recognition result and a vocabulary storage means which stores vocabulary items as the target to be recognized during said speech recognition, it is equipped with  
an alternative expression storage means which stores alternative expressions for replacing arbitrary expressions,

an alternative expression detection means which detects the same expressions as the aforementioned vocabulary items stored in the aforementioned alternative expression storage means according to the input speech information,

a substitution expression storage means which stores individual words obtained by further dividing the vocabulary stored in the aforementioned vocabulary storage means, and

a processing means which carries out the speech recognition of said non-replaced part of the input speech information other than the part where the aforementioned alternative expression is detected by the aforementioned alternative expression detection means while using the vocabulary items stored in the aforementioned alternative expression storage means as the target of the speech recognition and retrieves a vocabulary item appropriate for the replaced expression from the vocabulary items stored in the aforementioned substitution expression storage means using the result of said speech recognition.

5. The speech input interpreter described under Claim 4 characterized in that the aforementioned processing means carries out the aforementioned speech recognition by the unit of a syllable or a phoneme, detects a part where one of the aforementioned normal vocabulary items is added and uttered as a part of the alternative expression in reference to the result of said recognition by the unit of a syllable or a phoneme, and selects an expression which matches the aforementioned detection result with the highest priority when retrieving the replaced expression from the vocabulary items stored in the aforementioned substitution expression storage means.

6. The speech input interpreter described under Claim 4 characterized in that the aforementioned alternative expression detection means carries out metrical analysis of the input speech, and the aforementioned processing means selects an expression which matches or approximates the metrical requirement obtained as a result of the aforementioned analysis with the highest priority when retrieving the replaced expression from the vocabulary items stored in the aforementioned substitution expression storage means.

7. A speech input interpretation method characterized in that in a speech input interpretation method in which an input speech is interpreted, and information on applicable vocabulary items is output, speech recognition is carried out with respect to the input speech, an alternative expression is detected according to the aforementioned speech recognition result based on information pertaining to normal vocabulary items which consider the possibility that some of the prescribed normal vocabulary items may be replaced by prescribed alternative expressions during speech input, and

a query is run in the information pertaining to the prescribed normal vocabulary items based at least on the vocabulary items other than said alternative expression included in the aforementioned input speech recognition result when the aforementioned alternative expression is detected from the aforementioned recognition result in order to find an applicable vocabulary item.

8. The speech input interpretation method described under Claim 7 characterized in that when multiple applicable vocabulary items are retrieved, the priorities of the applicable vocabulary items are evaluated based at least on the phonemic characteristic of the speech corresponding to the aforementioned alternative expression.

9. A speech input interpretation method characterized in that in a speech input interpretation method in which input speech is interpreted, and information on applicable vocabulary items is output,

speech recognition and metrical analysis are carried out with respect to the speech input via a speech input device in reference to a vocabulary storage means which stores alternative expressions obtained by replacing some prescribed normal vocabulary items as the target of speech recognition by alternative expressions used to replace arbitrary words as one type of vocabulary, and

the aforementioned alternative expression part is substituted with the aforementioned normal vocabulary part based on the result of the aforementioned speech recognition of the aforementioned input speech and the result of the aforementioned metrical analysis in reference to the aforementioned metrical information storage means which stores information on the notations of and metrical information on the aforementioned normal vocabulary items which do not contain the aforementioned alternative expressions out of the vocabulary items stored in the aforementioned vocabulary storage means.

/3

10. A speech input interpretation method characterized in that in a speech input interpretation method in which an input speech is interpreted by means of speech recognition, and information on applicable vocabulary items in a vocabulary storage means which stores vocabulary items as the target to be recognized is output during said speech recognition, an expression identical to a vocabulary item stored in an alternative expression storage means which stores alternative expressions for replacing arbitrary expressions is detected from speech information input, and

the speech recognition of the non-replaced part of the input speech information other than the part where the aforementioned alternative expression is detected is carried out while using vocabulary items stored in a substitution expression storage means which stores individual words obtained by further dividing the vocabulary stored in the aforementioned vocabulary storage means as the target of the speech recognition in order to retrieve a vocabulary item appropriate

for the replaced expression from the vocabulary items stored in the aforementioned substitution expression storage means using the result of said speech recognition.

11. The speech input interpretation method described under Claim 10 characterized in that during the retrieval of the aforementioned vocabulary item, the aforementioned processing means carries out the aforementioned speech recognition by the unit of a syllable or a phoneme, detects a part where one of the aforementioned normal vocabulary items is added and uttered as a part of the alternative expression in reference to the result of said recognition by the unit of a syllable or a phoneme, and selects an expression which matches the aforementioned detection result with the highest priority when retrieving the replaced expression from the vocabulary items stored in the aforementioned substitution expression storage means.

12. The speech input interpretation method described under Claim 11 characterized in that an expression which matches or approximates the metrical requirement obtained as a result of the metrical analysis of the input speech is selected with the highest priority when retrieving the replaced expression from the vocabulary items stored in the aforementioned substitution expression storage means.

#### Detailed explanation of the invention

[0001]

##### Technical field of the invention

The present invention pertains to a speech input interpreter and a speech input interpretation method for interpreting input speech.

[0002]

##### Prior art

In the recent years, in the case of computer systems, including personal computers, audio information can be input in addition to input by means of the conventional mouse and keyboard.

[0003]

In addition, due to advances in natural language analysis and natural language generation, or speech recognition and speech synthesis technology, or interactive processing technology, there is an increasing demand for speech-based interactive systems which allow users to interact [with computers] by means of speech input/output; and a variety of speech-based interactive systems are being developed, for example, "TOSBURG-II," an interactive system utilizing speech input based on a natural speech (Japan Electronics and Information Technology Industries Association Theses, Vol. 1J77-D-11, No. 8, pp. 1417-1428, 1994).

[0004]

An input method utilized for this kind of speech-based interactive system can be handled by anyone in that it does not require any special skill with the keyboard, so that it is expected to be used for a social system to be utilized by everybody, resulting in an increasing demand for a higher level of speech processing technology.

[0005]

Conventionally, to interpret a speech input, the speech of a user is input through a microphone, a candidate speech analysis unit is estimated based on signal intensity, for example; a characteristic pattern is extracted through an analysis utilizing FFT (fast Fourier transformation) of analysis unit items; the extracted pattern is checked against a prepared standard pattern using a complex similarity method, DP (dynamic programming) method, or HMM (hidden Markov model), for example, in order to recognize the speech input; and a structural analysis and a semantic analysis are applied to the speech recognition result in order to extract the content of the speech input by the user and the intention of the speech.

[0006]

In the past, to carry out speech recognition of the speech input interpretation method of said speech-based interactive system, matching with a word or sentence pattern prepared in advance was performed. However, with said method, a burden was placed upon a user in that the user needed to memorize words or sentences which could be spoken (that is, words or sentences said system was able to interpret) accurately.

[0007]

Furthermore, when the user memorized only some of the words or sentences which could be spoken and input said partially memorized part, it was misrecognized as it was considered a speech input different from the prepared pattern. As a result, an operation contradictory to the intention of the user was output in many cases, imposing a burden on the user.

[0008]

A system for a road guidance task, for example, will be explained as a specific example of a social system. When the only information a user has is the portion "Tokyo Hotel" of "Tokyo Stayin [transliteration] Hotel," and "Tokyo something Hotel" is input in an attempt to extract information pertaining to said hotel, because it is different from name patterns of actually

existing hotels prepared in the system in advance, a recognition error occurs, and information contradictory to the intention of the user is presented, providing no service to the user at all.

[0009]

In addition, if only the rhythms of words or sentences which the user could utter (or those expected to be registered in the system without a doubt) were stored, when another word or a sentence which had the same rhythm as that of said word or the sentence was input, the conventional system could not accept it as a proper input, resulting in a recognition error. Thus, the operation intended by the user was never carried out, resulting in a burden on the user.

[0010]

Similarly, using the system for a road guidance task as a specific example of a social system, assume that when a user attempts to obtain information pertaining to "Marunoguchi Hotel," the rhythm of "Marunoguchi Hotel" and the partial information "... Hotel" are the only information said user has. In such case, even when the user utters "lalalalala Hotel" or "honyalalala\* Hotel" or "talalalala Hotel" to mean "something Hotel" using the rhythm of "Marunoguchi Hotel" carefully (or by imitating it) in an attempt to extract information pertaining to said hotel, a recognition error occurs, and information contradictory to the intention of the user is presented, providing no service to the user at all.

/4

[0011]

As described above, with the conventional speech input interpretation method, because only those patterns of words or sentences prepared in advance could be understood, a great burden was imposed on users.

[0012]

Problems to be solved by the invention

As described above, when the conventional speech input interpretation method is applied to a device involving a speech input, because word or sentence patterns which can be accepted as a speech input are restricted to those registered in advance, a user needs to memorize the sentences he/she can utter accurately, resulting in the problem of an increasing burden on the user.

---

\* [Translator's note: Transliteration of a nonsense word.]



[0013]

In addition, even when the user remembers only part of a word or a sentence he/she can utter and has input said part he/she remembers, it is considered to be a speech input different from the prepared pattern, resulting in a recognition error. As a result, an operation contradictory to the intention of the user was output in many cases, resulting in an increased burden on the user.

[0014]

In addition, when the user remembers only the rhythm of a word or a sentence he/she can utter, the conventional system cannot accept it as a proper input, a recognition error occurs, and the operation intended by the user is never carried out, resulting in an increased burden on the user.

[0015]

The present invention was made in consideration of the aforementioned situation, and its purpose is to present a speech input interpreter capable of interpreting [a speech input] correctly in order for an application part to operate properly without requiring the user to memorize words and sentences he/she can utter accurately.

[0016]

In addition, the present invention aims to present a speech input interpreter capable of restraining interpretation errors in order to direct an output of the speech input device toward the user's intention even when the user remembers only part of a word or a sentence he/she is allowed to utter.

[0017]

In addition, the present invention aims to present a speech input interpreter and a speech input interpretation method by which interpretation errors can be restrained in order to direct an output of the speech input device toward the user's intention even when the user remembers only part or the rhythm of a word or a sentence he/she is allowed to utter.

[0018]

Means to solve the problems

The present invention (Claim 1) is characterized in that in a speech input interpreter which interprets input speech and outputs information on applicable vocabulary items, it is equipped with a means which stores first information pertaining to normal vocabulary items and second information which considers the possibility that some of said normal vocabulary items

may be replaced by prescribed alternative expressions during speech input, a means for speech recognition of the input speech, a means which detects the aforementioned alternative expression from the result of the aforementioned speech recognition based on the aforementioned second information, and a means which runs a query in the aforementioned first information based at least on the vocabulary items other than said alternative expression included in the aforementioned input speech recognition result when the aforementioned alternative expression is detected from the aforementioned recognition result by said means in order to find said applicable vocabulary item.

[0019]

Preferably, a means which evaluates the priorities of applicable vocabulary items based at least on the phonemic characteristic of the speech corresponding to the aforementioned alternative expression when multiple applicable vocabulary items are retrieved may be further provided.

[0020]

The present invention (Claim 3) is characterized in that in a speech input interpreter which interprets an input speech and outputs information on applicable vocabulary items, it is equipped with a vocabulary storage means which stores alternative expressions obtained by replacing some of prescribed normal vocabulary items as the target of the speech recognition by alternative expressions used to replace arbitrary words as one type of vocabulary item, a metrical information storage means which stores information on the notations of and metrical information on the aforementioned normal vocabulary items which do not contain the aforementioned alternative expressions out of the vocabulary items stored in the aforementioned vocabulary storage means, a speech analysis means which carries out speech recognition and metrical analysis of a speech input through a speech input device in reference to the aforementioned vocabulary storage means, and a substitution expression verification means which replaces the aforementioned replaced part by the aforementioned normal vocabulary item based on the result of the aforementioned speech recognition of the aforementioned input speech and the result of the aforementioned metrical analysis by the aforementioned speech analysis means in reference to the aforementioned metrical information storage means.

[0021]

With the present invention, even when the user does not remember the vocabulary items stored in the vocabulary storage means clearly, an unclear part can be input through speech using an alternative expression, and an appropriate expression corresponding to the alternative

expression input is retrieved so as to substitute it with an appropriate vocabulary item which does not contain the alternative expression.

[0022]

The present invention (Claim 4) is characterized in that in a speech input interpreter equipped with a means which analyzes a speech input from a speech input device for speech recognition and outputs a speech analysis result including a speech recognition result and a vocabulary storage means which stores vocabulary items as the target to be recognized during said speech recognition, it is equipped with an alternative expression storage means which stores alternative expressions for replacing arbitrary expressions, an alternative expression detection means which detects the same expressions as the aforementioned vocabulary items stored in the aforementioned alternative expression storage means according to speech information input, a substitution expression storage means which stores individual words obtained by further dividing the vocabulary stored in the aforementioned vocabulary storage means, and a processing means which carries out the speech recognition of said non-replaced part of the input speech information other than the part where the aforementioned alternative expression is detected by the aforementioned alternative expression detection means while using the vocabulary items stored in the aforementioned alternative expression storage means as the target of the speech recognition and retrieves a vocabulary item appropriate for the replaced expression from the vocabulary items stored in the aforementioned substitution expression storage means using the result of said speech recognition.

/5

[0023]

With the present invention, even when the user does not remember the vocabulary items stored in the vocabulary storage means clearly, an unclear part can be input through speech using an alternative expression, and an expression which replaces an arbitrary expression can be detected from the speech input so as to retrieve an appropriate expression corresponding to the alternative expression detected.

[0024]

Preferably, the aforementioned processing means may be configured as such that it carries out the aforementioned speech recognition by the unit of a syllable or a phoneme, detects a part where one of the aforementioned normal vocabulary items is added and uttered as a part of the alternative expression in reference to the result of said recognition by the unit of a syllable or a phoneme, and selects an expression which matches the aforementioned detection result with the

highest priority when retrieving the replaced expression from the vocabulary items stored in the aforementioned substitution expression storage means.

[0025]

As a result, a more appropriate expression adaptable to information on a partially correct utterance in a speech input containing the partially correct utterance within the alternative expression used by the user can be retrieved.

[0026]

Preferably, the aforementioned alternative expression detection means may be configured as such that it carries out metrical analysis of the input speech, and the aforementioned processing means selects an expression which matches or approximates the metrical requirement obtained as a result of the aforementioned analysis with the highest priority when retrieving the replaced expression from the vocabulary items stored in the aforementioned substitution expression storage means.

[0027]

The present invention (Claim 7) is characterized in that in a speech input interpretation method in which an input speech is interpreted, and information on applicable vocabulary items is output, speech recognition is carried out with respect to the input speech, an alternative expression is detected according to the aforementioned speech recognition result based on information pertaining to normal vocabulary items which take into consideration of the possibility that some of the prescribed normal vocabulary items may be replaced by prescribed alternative expressions during the speech input, and a query is run in the information pertaining to the prescribed normal vocabulary items based at least on the vocabulary items other than said alternative expression included in the aforementioned input speech recognition result when the aforementioned alternative expression is detected from the aforementioned recognition result in order to find an applicable vocabulary item.

[0028]

Preferably, [the method] may be configured as such that when multiple applicable vocabulary items are retrieved, the priorities of the applicable vocabulary items are evaluated based at least on the phonemic characteristic of the speech corresponding to the aforementioned alternative expression.

[0029]

The present invention (Claim 9) is characterized in that in a speech input interpretation method in which an input speech is interpreted, and information on applicable vocabulary items is output, speech recognition and metrical analysis of the speech are carried out with respect to the speech input via a speech input device in reference to a vocabulary storage means which stores alternative expressions obtained by replacing some of prescribed normal vocabulary items as the target of speech recognition by alternative expressions used to replace arbitrary words as one type of vocabulary, and the aforementioned alternative expression part is substituted with the aforementioned normal vocabulary part based on the result of the aforementioned speech recognition of the aforementioned input speech and the result of the aforementioned metrical analysis in reference to the aforementioned metrical information storage means which stores information on the notations of and metrical information on the aforementioned normal vocabulary items which do not contain the aforementioned alternative expressions out of the vocabulary items stored in the aforementioned vocabulary storage means.

[0030]

The present invention (Claim 10) is characterized in that in a speech input interpretation method in which an input speech is interpreted by means of speech recognition, and information on applicable vocabulary items in a vocabulary storage means which stores vocabulary items as the target to be recognized is output during said speech recognition, an expression identical to a vocabulary item stored in an alternative expression storage means which stores alternative expressions for replacing arbitrary expressions is detected from input speech information, and the speech recognition of the non-replaced part of the input speech information, other than the part where the aforementioned alternative expression is detected, is carried out while using vocabulary items stored in a substitution expression storage means which stores individual words obtained by further dividing the vocabulary stored in the aforementioned vocabulary storage means as the target of the speech recognition in order to retrieve a vocabulary item appropriate for the replaced expression from the vocabulary items stored in the aforementioned substitution expression storage means using the result of said speech recognition.

[0031]

Preferably, [the method] may be configured as such that during the retrieval of the aforementioned vocabulary item, the aforementioned processing means carries out the aforementioned speech recognition by the unit of a syllable or a phoneme, detects a part where one of the aforementioned normal vocabulary items is added and uttered as a part of the alternative expression in reference to the result of said recognition by the unit of a syllable or a

phoneme, and selects an expression which matches the aforementioned detection result with the highest priority when retrieving the replaced expression from the vocabulary items stored in the aforementioned substitution expression storage means.

[0032]

Preferably, [the method] may be configured as such that an expression which matches or approximates the metrical requirement obtained as a result of the metrical analysis of the input speech is selected with the highest priority when retrieving the replaced expression from the vocabulary items stored in the aforementioned substitution expression storage means.

[0033]

With the present invention, when a function to detect a wildcard expression as a replacement for a clear expression and a function to retrieve said replaced appropriate expression for substitution are added, or when a speech analysis function involving a vocabulary storage means containing a vocabulary item actually substituted with the wildcard expression and a function to retrieve said replaced appropriate expression for substitution are added, even when the user remembers only some of vocabulary items he/she is allowed to utter, a speech input containing the wildcard expression can be accepted in order to interpret said speech input.

[0034]

In addition, with the present invention, even when the user remembers only the rhythm of a vocabulary item he/she is allowed to utter, a speech input containing a wildcard expression corresponding to it can be accepted so as to interpret said speech input.

[0035]

As such, the present invention offers a great practical effect in that a flexible speech input interpreter capable of accepting and interpreting a speech input even when a user does not clearly remember a vocabulary item a speech input device can accept. /6

[0036]

Embodiments of the invention

Embodiments of the present invention will be explained below with reference to Figures.

[0037]

First embodiment

First, a first embodiment of the present invention will be explained.

[0038]

An example configuration of the speech input interpreter pertaining to the present invention is shown in Figure 1. As shown in Figure 1, speech input interpreter 1 of the present embodiment is equipped with speech analysis part 101, vocabulary storage part 102, substitution expression verification part 103, and metrical information storage part 104. Furthermore, an A/D converter for converting input speech from an analog signal to a digital signal may be provided either inside of speech input interpreter 1 or on the side of speech input device 100.

[0039]

Speech analysis part 101 is connected to substitution expression verification part 103, vocabulary storage part 102, and speech input device 100, such as a microphone; whereby, speech recognition of continuous words is carried out while targeting the vocabulary stored in vocabulary storage part 102 using a system of the kind disclosed in "Algorithms for Speech Recognition of Continuous Words and Continuous Syllables Based on a Pattern Matching Method" (Japan Electronics and Information Technology Industries Association Theses, Vol. J-66-D, [No.] 6, pp. 637-644), for example. Furthermore, speech is analyzed based on information on its pitch pattern in order to generate a metrical parameter using a system of the kind disclosed in "Keyword Spotting Utilizing Pitch Pattern Information" (Collection of lectures by the Acoustical Society of Japan, September Heisei 8[1996], pp. 29-30), for example. Then, the information shown in Figure 4 is given to substitution expression verification part 103. Here, the system for speech recognition of continuous words and the method for generating the metrical parameter are not restricted to those described above, and other systems may also be utilized

[0040]

Vocabulary storage part 102 as a part for storing vocabulary items as the target of speech recognition is connected to speech analysis part 101. It stores information of the kind shown in Figure 2 with respect to each normal vocabulary item (in this case, there is no information pertaining to any wildcard expression in Figure 2). In addition, it stores information of the kind shown in Figure 2 regarding each normal vocabulary item obtained by substituting part of the normal vocabulary item with a wildcard expression, that is, an expression substituted with several arbitrary words such as "something" or "honyalala."

[0041]

While the details of the information in Figure 2 will be described later, the notation of format of “representation” information will be discussed first. Because a recognition result cannot be expressed as a series of multiple words during the speech recognition of continuous words by speech analysis part 101, the divisions between said words are indicated using the symbol “/” (slash). In addition, the symbol “/” will be used to indicate the divisions between said words in the following explanation, too.

[0042]

In addition, either multiword substitution words, such as “something,” which are considered to substitute several words or rhythmic words, such as “honyalala,” which are considered to indicate the rhythms of expressions to be substituted, or both, are defined as wildcard expressions to be used. Specific contents and kinds of the multiword substitution words and the rhythmic words to be used may be decided according to the system as needed.

[0043]

An example vocabulary generated from “Tokyo Stayin Hotel,” the wildcard expression “something” as a multiword substitution word, and the rhythmic word “honyalala” is shown in Figure 3. Here, it is clear that vocabulary items obtained by substituting “Tokyo,” “Stayin,” and/or “Hotel” with the wildcard expression are generated, and particularly that the rhythmic word “honyalala” is extended to the same length as that of the expression substituted (in this case, the length is adjusted by the number of “la”).

[0044]

Figure 2 is a list of information stored in vocabulary storage part 102. A case involving the vocabulary item “Tokyo honyalala Hotel” is shown therein. “Representation” information refers to information indicating the character string for said vocabulary item. In the case of the example in Figure 2, it is stored as a representation comprising 3 continuous words as in “Tokyo/honyalala/Hotel.” “Presence/absence of wildcard expression” information refers to information which indicates whether the aforementioned wildcard expression is included in said vocabulary item or not. In this case, “honyalala” corresponds to the wildcard expression, and “Yes” is stored accordingly. “Expression type” information refers to information which indicates whether each of the words contained in said vocabulary item is a wildcard expression or a non-wildcard expression. “Replace” is given to a word of wildcard expression, and “OK” is given to a non-wildcard expression. In said example, because the words “Tokyo” and “Hotel” are non-wildcard expressions, and “honyalala” is a wildcard expression, (OK/Replace/OK) is given.



“Wildcard expression type” information refers to information which indicates the wildcard expression contained in a given vocabulary item is a multiword substitution word or information representing a rhythmic word. In this case, because “honyalala” is defined as a rhythmic word, “Rhythmic word” is stored. “Speech recognition parameter” information is used to describe the necessary parameters used by speech analysis part 101 for speech recognition (since the speech recognition method used here is not intrinsic to the present invention, a detailed explanation of said parameters will be omitted).

[0045]

Figure 4 shows a list of information given from speech analysis part 101 to substitution expression verification part 103. The case involving the vocabulary item “Tokyo honyalala Hotel” is also shown. “Recognition result” information refers to information indicating the result a representation for the continuous word recognized by speech analysis part 101. In the case of the example in Figure 4, “Tokyo/honyalala/Hotel” is shown as the recognition result of the speech signal input. “Word utterance time” information refers to information indicating the utterance time of each word obtained during the continuous word recognition by speech analysis part 101. (650 msec/820 msec/510 msec) is shown in said example, and these numbers indicate the utterance times corresponding to “Tokyo,” “honyalala,” “Hotel,” respectively. “Metrical parameter” information refers to information indicating the metrical parameter analyzed by speech analysis part 101. While said information takes different forms depending on the metrical parameter analysis means used, a case in which changes in the intonation or the basic frequency over time is used is shown here. In this case, a would-be-obtained parameter is shown schematically. The arrow symbol “→” used in Figure 4 indicates the intonation of said expression schematically; wherein, arrows at the top indicates the high portion of the intonation, and arrows at the bottom indicate the low portions of the intonation. “Presence/absence of wildcard expression” information refers to information which indicates whether the aforementioned wildcard expression was included in the input speech or not. In this case, because “honyalala” corresponds to the wildcard expression, “Yes” is output. “Expression type” information refers to information which indicates whether each of the words contained in said vocabulary item is a wildcard expression or a non-wildcard expression. Said information can be obtained in reference to “expression type” information in Figure 2 pertaining to the applicable vocabulary item, and its notation method is identical to that of “expression type” information in Figure 2. In the case of the example in Figure 4, because “Tokyo” and “Hotel” are non-wildcard expressions, and “honyalala” is a wildcard expression, the information is given as (OK/Replace/OK). “Presence/absence of wildcard expression” information refers to information

used to identify whether a wildcard expression is a multiword substitution word or a rhythmic word. In this case, because “honyalala” is a rhythmic word, “Rhythmic word” is output.

[0046]

Substitution expression verification part 103 is connected to speech analysis part 101 and metrical information storage part 104; whereby, when a wildcard expression is detected, it finds an appropriate expression corresponding to said wildcard expression. Details of said part will be described later.

[0047]

Metrical information storage part 104 is connected to substitution expression verification part 103; whereby, it stores information of the kind shown in Figure 5 pertaining to vocabulary items not containing any wildcard expression out of the vocabulary items stored in vocabulary storage part 102.

[0048]

Figure 5 is a list of information stored in metrical information storage part 104. A case involving “Tokyo honyalala Hotel” is shown therein. “Representation” information refers to information indicating the representation of said vocabulary item. “Standard time” information indicates the utterance time of the sample expression stored. If said vocabulary item comprises continuous words and can thus be divided, utterance time for each word is stored. Notation method of said information is identical to that of “word utterance time” information in Figure 4. “Metrical” information indicates metrical information obtained by analyzing the sample expression stored. However, the method used to analyze the metrical information must be identical to the method used by speech analysis part 101. In addition, the metrical information output from metrical information storage part 104 must be in the same format as that used for the metrical parameter information given from speech analysis part 101 to substitution expression verification part 103. Like Figure 4, metrical information which would be obtained after the analysis is shown schematically in the example in Figure 5.

[0049]

Figure 6 is a flowchart of the operations of substitution expression verification part 103 which plays an important role in the present embodiment. The processing flow will be explained below with reference to Figure 6.

[0050]

Step S101

Here, whether the speech recognition result by speech analysis part 101 contains a wildcard expression or not is checked. It can be checked using the "presence/absence of wildcard expression" information in Figure 4 given from speech analysis part 101. If a wildcard expression is present, [it advances] to Step S102; and if no wildcard expression is present, the recognition result is output, and the processing is ended.

[0051]

Step S102

In this step, vocabulary items which match the speech recognition result and can be output are selected from metrical information storage part 104. For example, a non-wildcard expression part contained in the speech recognition result is obtained in reference to the expression type information in the information (Figure 4) given from speech analysis part 101 utilizing the representation information of the information (Figure 5) stored in metrical information storage part 104, and vocabulary items which match the condition pertaining to the position where said non-wildcard expression is present are selected while using the non-wildcard expression part as a condition based on an assumption that a wildcard expression has a length longer than one word.

[0052]

For example, assuming that the obtained speech recognition result is "Tokyo/something/Hotel," and the expression type is "OK/Replace/OK," vocabulary items involving the word "Tokyo" at the beginning, the word "Hotel" at the end, and at least 1 word between them can be selected as matching vocabulary items; that is, "Tokyo/Stayin/Hotel," "Tokyo/Inter\*/Continental/Hotel," and so forth.

[0053]

Step S103

In this step, whether or not the wildcard expression contained in the speech recognition result is a multiword substitution word or a rhythmic word is identified. This can be checked using the "wildcard expression type" information in Figure 4 given from speech analysis part 101. If it is a rhythmic word, [it advances] to Step S104; and if it is a multiword substitution word, the vocabulary item extracted in Step S102 is output, and the processing is ended. Furthermore, if

---

\* [Editor's note: For the purposes of this patent, "inter" will be considered to be a separate word for "continental."]

there are multiple vocabulary items to be output, either some or all of them may be output. This processing, that is, multiple answers being presented to let a user make a selection, is decided according to processing (200 in the figure) specific to the application at the output destination.

[0054]

#### Step S104

In this step, the “word utterance time” information given from speech analysis part 101 and the “standard time” information stored in metrical information storage part 104 are compared in order to further restrict the vocabulary items to be output among the vocabulary items extracted in Step S102. For example, in the case of “Tokyo/honyalala/Hotel,” respective ratios between the utterance times for the non-wildcard expression parts “Tokyo” and “Hotel” and the standard times for “Tokyo” and “Hotel” in the standard time information pertaining to the target vocabulary items stored in metrical information storage part 104 are calculated, the utterance time for the wildcard part in the input signal is extended using the mean value of said ratios, the extended utterance time and the standard time are compared, and only those within a given threshold are extracted. Furthermore, if a vocabulary item is not specified through said processing, priority order of the vocabulary items to be output can be decided by comparing their times.

/8

[0055]

#### Step S105

In this step, a vocabulary item to be output out of the vocabulary items extracted in Step S104 is decided by comparing the “metrical parameter ” information given from speech analysis part 101 with the “metrical parameter ” information stored in metrical information storage part 104. For example, DP method based matching is carried out using the method disclosed in “Keyword Spotting Utilizing Pitch Pattern Information” (collection of lectures by the Acoustical Society of Japan, September Heisei 8[1996], pp. 29-30). Furthermore, although said comparison method differs depending on the metrical parameter configured, an arbitrary metrical comparison method may be utilized in the present embodiment as long as the configured parameter can be utilized. Then, a vocabulary item with the most closely resembling metrical information to that of the speech uttered is output before ending the processing. Alternatively, if there are multiple candidates, they may be prioritized for output in the order of the resemblance of their metrical information.

[0056]

Above are the configuration and the function of substitution expression verification part 103 and the processing method pertaining to the present invention.

[0057]

Next, the aforementioned speech input interpretation method will be explained in further detail below. Here, a specific application example of the operations of a map information system when speech is input by a user will be explained.

[0058]

Assume that information on 4 hotels (Palace Hotel, Tokyo Stayin Hotel, Tokyo Marunoguchi hotel, and Tokyo Inter Continental Hotel) is registered to said map information system, and the names of the 4 hotels are stored in vocabulary storage part 102. In addition, assume that the aforementioned rhythmic word “honyalala” is registered to vocabulary storage part 102 as a wildcard expression, and the vocabulary items shown in Figure 7 are registered to vocabulary storage part 102 together with vocabulary items generated from the aforementioned 4 hotels and “honyalala.”

[0059]

Also, assume that representation information, metrical information, and standard time information are obtained from the registered 4 hotel names, and information of the kind shown in Figure 8 is stored in metrical information storage part 104.

[0060]

Then, assume that a user wants to make an inquiry regarding “Tokyo Stayin Hotel,” but he/she does not remember the part “Stayin” clearly, and enters the speech input “Tokyo honyalala Hotel.” Here, assume that the wildcard expression “honyalala” is uttered with the rhythm of “Stayin” intentionally.

[0061]

Operations of the respective parts in the case of the present specific example will be described.

[0062]

First, speech analysis part 101 recognizes the continuous word using the vocabulary items in Figure 7 with respect to the speech input. Then, “Tokyo/honyalala/Hotel” is selected as the

result of the recognition, and the information shown in Figure 9 is output to substitution expression verification part 103 together with utterance time information obtained during the recognition processing and metrical information extracted from the input speech.

[0063]

Upon receiving said information, substitution expression verification part 103 carries out the following processing.

[0064]

Step S101

It judges that the recognition result contains a wildcard expression based on the “presence/absence of wildcard expression” information received and advances to Step S102.

[0065]

Step S102

It decides that non-wildcard expressions are “Tokyo” and “Hotel” based on the recognition result information “Tokyo/honyalala/Hotel” and the expression type information (OK/Replace/OK) and retrieves those which match the condition pertaining to the positions of these 2 words from the vocabulary items (Figure 8) registered to metrical information storage part 104. In this case, vocabulary items containing “Tokyo” at the beginning, “Hotel” at the end, and at least 1 word between them are considered to match the retrieval condition. As such, “Tokyo Stayin Hotel,” “Tokyo Marunoguchi Hotel,” and “Tokyo Inter Continental Hotel” are retrieved; and “Palace Hotel” is either excluded as a candidate to be output or considered as a candidate with a lower priority.

[0066]

Step S103

It decides that the wildcard expression “honyalala” is a rhythmic word based on the wildcard expression type information received and advances to Step S104.

[0067]

Step S104

Of the vocabulary items selected in Step S102, the standard time information (Figure 8) on “Tokyo Stayin Hotel” and the word utterance time information given from speech analysis part 101 are compared first. For example, ratios (standard time information/word utterance time information) are calculated with respect to the wildcard expressions “Tokyo” and “Hotel,”

“Tokyo”:  $700/650 = 1.0769$  and “Hotel”:  $550/510 = 1.0784$  results. The average of these ratios is calculated, and the numeric value (1.0777) obtained as a result is used as an expansion factor for handling the input time and the standard time in metrical information storage part 104 on the same scale. As such, after the “honyalala” part corresponding to the wildcard expression is extended, the input time for the wildcard expression part becomes  $820 \text{ msec} \times 1.0777 = 884 \text{ msec}$ . Next, the standard time for the “Stayin” part, which is located between “Tokyo” and “Hotel” and considered to have been replaced by the wildcard expression, becomes 900 msec. Then, these 2 input times are compared (for example, threshold processing) in order to check whether the times for the wildcard expression part match or not. Results of the aforementioned calculation pertaining to the vocabulary items selected in Step S102 are shown in Figure 10.

/9

[0068]

Here, because “Inter/Continental” part of “Tokyo Inter Continental Hotel” is considered to have been replaced by “honyalala,”  $650 + 1050 = 1700 \text{ msec}$  corresponding to “Inter” and “Continental” out of the standard time ( $700 \text{ msec}/650 \text{ msec}/1050 \text{ msec}/550 \text{ msec}$ ) is considered the standard time corresponding to the wildcard expression “honyalala.” Then, the difference between the time after the expansion and the standard time is calculated, and those whose absolute value is greater than a given threshold are excluded as candidates to be output. Assuming that said threshold is 100 msec, “Tokyo Inter Continental Hotel” is either excluded as a candidate to be output or considered as a candidate with a lower priority.

[0069]

Step S105

Metrical information matching is applied to “Tokyo Stayin Hotel” and “Tokyo Marunoguchi Hotel” which have not been excluded during the aforementioned processing. Then, either a vocabulary item with metrical information close to the metrical information given from speech analysis part 101 or the vocabulary item with the higher priority is output as a result. Here, the metrical information on “Tokyo Stayin Hotel” is judged to be closer to the rhythm of the input speech, “Tokyo Stayin Hotel” is output as the vocabulary item with the higher priority, and an appropriate processing (200 in the figure) specific to the application of the output destination is carried out. In addition, if the application-specific processing can be applied to multiple candidates, “Tokyo Marunoguchi Hotel” is output as a lower-priority candidate, and “Tokyo Inter Continental Hotel” and “Palace Hotel” are also output in said order as even lower-priority candidates if necessary.

[0070]

Now, the processing to be carried out when “Tokyo honyalala Hotel” is input in the form of a speech is completed.

[0071]

It is clear from the aforementioned explanation that with the speech input interpreter pertaining to the present embodiment, even when a user does not remember the name “Tokyo Stayin Hotel” clearly, he/she can enter the speech input of “Tokyo honyalala Hotel” using the wildcard expression for the unclear part in order to have information on appropriately interpreted name output to the application part, and that the speech information the user input is utilized more effectively in that the user can input the rhythm which he/she knows but cannot be expressed by characters in the form of “Tokyo honyalala Hotel” using the wildcard expression in order for the system pertaining to the present embodiment to interpret the utterance time information and rhythmic information so as to give the priority to “Tokyo Stayin Hotel” over “Tokyo Inter continental Hotel” and “Tokyo Marunoguchi Hotel” which are names of the same format, that is, “Tokyo ... Hotel.”

[0072]

Second embodiment

Next, a second embodiment of the present invention will be explained.

[0073]

Although the speech recognition system utilized continuous word recognition in the first embodiment, the speech recognition system can be applied even when the continuous word recognition is not involved in the present embodiment.

[0074]

An example configuration of the speech input interpreter pertaining to the present embodiment is shown in Figure 11. As shown in Figure 11, speech input interpreter 2 of the present embodiment is equipped with speech analysis part 201, vocabulary storage part 202, wildcard expression detection part 203, wildcard expression storage part 204, substitution expression verification part 205, and substitution expression storage part 206. Furthermore, an A/D converter for converting an input speech from an analog signal into a digital signal may be provided either inside of speech input interpreter 2 or on the side of speech input device 100.



[0075]

Speech analysis part 201 is connected to substitution expression verification part 205, vocabulary storage part 202, and substitution expression storage part 206; whereby, words in a speech are recognized using vocabulary items either in vocabulary storage part 202 or substitution expression storage part 206 whichever is specified as a speech recognition request arrived from substitution expression verification part 205, and the results are output to substitution expression verification part 205. In addition, individual syllables are recognized as requested by the recognition method, and the results are output to substitution expression verification part 205 in the form of a Mauler code string. Here, since the speech recognition method used here is not intrinsic to the present invention, their detailed explanation will be omitted.

[0076]

Vocabulary storage part 202 is a part connected to speech analysis part 201 and substitution expression verification part 205 in order to store (normal) vocabulary items as the target of speech recognition; whereby, it stores the information shown in Figure 12 regarding the respective vocabulary items as the target of speech recognition in a format which can be referred to and utilized by speech analysis part 201 and substitution expression verification part 205.

[0077]

Figure 12 is a list of information stored in vocabulary storage part 202. Information to be stored in vocabulary storage part 202 regarding “Tokyo Stayin Hotel” is shown as an example. “Representation character string” information refers to a character string representing vocabulary items to be registered. “Mauler code string” information refers to the reading of the representation character string expressed in the form of a string of Mauler codes. “Mauler code string” information indicates the number of Mauler codes in the Mauler code string stored as Mauler code string information. Pitch pattern information on the speech is analyzed using the method disclosed in “Keyword Spotting Utilizing Pitch Pattern Information” (collection of lectures by the Acoustical Society of Japan, September Heisei 8[1996], pp. 29-30), for example, and a metrical parameter configured is stored as “metrical parameter ” information. Furthermore, the method for generating the metrical parameter is not restricted to the aforementioned system, and other systems may also be utilized. In addition, in the case of the example in Figure 12, would-be-obtained metrical information is shown schematically. This notation method is identical to that in the first embodiment. “Parameter required for speech recognition” information is used to describe a parameter necessary for speech recognition by speech analysis part 201 as needed for the implementation of the present invention (furthermore, since the speech

recognition method used here is not intrinsic to the present invention, detailed explanation of said parameter will be omitted here).

/10

[0078]

Wildcard expression storage part 204 is connected to wildcard expression detection part 203; whereby, it stores a wildcard expression as an expression substituted with several arbitrary words, such as "something" or "honyalala," in a format which can be referred to and utilized by wildcard expression detection part 203. Also, it divides the wildcard expressions to be stored into multiword substitution words, such as "something" and "so and so," comprising several words used for substitution and rhythmic words, such as "honyalala" and "talalala," which indicate the rhythm of an expression to be substituted when storing them.

[0079]

Wildcard expression detection part 203 is connected to speech input device 100, such as a microphone, wildcard expression storage part 204, and substitution expression verification part 205; whereby, it detects vocabulary items for the wildcard expressions stored in wildcard expression storage part 204 using the method disclosed in "Noise Immunity Learning During Speech Recognition Through Word Spotting" (collection of lectures by the Acoustical Society of Japan, Vol. J-74-D-II, September 1991, pp. 121-129), for example. Furthermore, the aforementioned system does not impose any restriction, and other detection systems may also be utilized as long as a specific vocabulary item can be detected. Then, wildcard expression detection part 203 gives the processing to substitution expression verification part 205 by giving it the information shown in Figure 13.

[0080]

Figure 13 is a list of information given from wildcard expression detection part 203 to substitution expression verification part 205. A case involving the vocabulary item "Tokyo honyalala Hotel" is shown therein. "Presence/absence of wildcard expression" information refers to information which indicates whether or not a wildcard expression is detected by wildcard expression detection part 203. In this example, "honyalala" corresponds to the wildcard expression, and "Yes" is output accordingly. "Original signal" refers to the original signal for the speech input; when a wildcard expression is detected, it is separated at the part where said wildcard expression is present and given to substitution expression verification part 205. In the example, the input "Tokyo honyalala Hotel" is separated by the wildcard expression "honyalala" into three parts, namely, "Tokyo," "honyalala," and "Hotel," and given to substitution expression verification part 205 in said order. When a wildcard expression is present, "wildcard expression

position” information indicates which of the separated [parts] of the original signal is the wildcard expression in the form of a numeric value. In this example, “honyalala” is found in the 2nd position of the three parts obtained by separating the signal, and 2 is output accordingly. “Wildcard expression type” refers to information which indicates whether the detected wildcard expression is a multiword substitution word or a rhythmic word. In this example, “honyalala” is considered as a rhythmic word. This differs depending on the information registered to wildcard expression storage part 204. “Mauler code string length of wildcard expression” refers to information which indicates the number of Mauler codes used for the detected wildcard expression when it is a rhythmic word. In this example, the wildcard expression “honyalala” has 4 as the number of Mauler codes. “Metrical information on wildcard expression” refers to information which indicates the rhythm of the detected wildcard expression when it is a rhythmic word. It is analyzed based on the pitch pattern information of the speech input and given to substitution expression verification part 205. Here, the method used to generate the metrical parameter must be such that the metrical parameter generated is in the same format as that used for storing in vocabulary storage part 202 or substitution expression storage part 206.

[0081]

Substitution expression verification part 205 is connected to wildcard expression detection part 203, speech analysis part 201, vocabulary storage part 202, and substitution expression storage part 206; whereby, when a wildcard expression is detected, it verifies an appropriate expression corresponding to said wildcard expression part. Details of said part will be described later.

[0082]

Substitution expression storage part 206 is connected to speech analysis part 201 and substitution expression verification part 205; whereby, it stores words, such as “Tokyo,” “Stayin,” and “Hotel,” further generated as meaningful words from the vocabulary item, for example, from “Tokyo Stayin Hotel,” stored in vocabulary storage part 202 or words which are used to create word combinations, such as “Tokyo Stayin” and “Stayin Hotel,” using the same format (Figure 12) as that used for vocabulary storage part 202.

[0083]

Figure 14 shows an outline of the operations of substitution expression verification part 205 which plays an important role in the present embodiment. The processing flow will be explained below with reference to Figure 14.

[0084]

Step S201

Whether or not the speech input contains a wildcard expression or not is checked. This can be checked using the “presence/absence of wildcard expression” information (Figure 13) given by wildcard expression detection part 203. If a wildcard expression is present, [it advances] to Step S204, or to Step S202 if no wildcard expression is present.

[0085]

Step S202

When a judgment that there is no wildcard expression is made, speech recognition processing is carried out as it is. Speech analysis part 201 is requested to carry out word recognition with respect to the original signal input using the vocabulary items stored in vocabulary storage part 202.

[0086]

Step S203

The speech recognition result output from speech analysis part 201 is given to application-specific processing (200 in the figure) or speech analysis processing of a higher level before ending the processing.

[0087]

Step S204

When a judgment that there is a wildcard expression is made, how the non-wildcard expression parts are uttered and input is checked. For example, a signal for the wildcard expression is obtained based on [the part] separated from the original signal and given by wildcard expression detection part 203 and the “wildcard expression position” information, and speech analysis part 201 is requested to carry out word recognition using the vocabulary items stored in substitution expression storage part 206 with respect to the signals for the non-wildcard expression parts (non-wildcard expression parts). The speech recognition result obtained in this step will be referred to as a “partial recognition result” in the following explanation. In addition, when there is no appropriate vocabulary item in substitution expression storage part 206, a decision that there is no partial recognition result corresponding to said non-wildcard expression parts is made.

[0088]

Step S205

Here, whether or not some kind of information is present between the non-wildcard expression parts and the wildcard expression part is checked. The purpose of doing so is to handle a case in which an utterance contains [a part] added before or after a wildcard expression as it is evident in “‘Su’\* something” when the user knows only the starting point and the ending point partially without having clear knowledge of the word pronunciations. When the utterance is input in the form of “‘Su’ something,” the wildcard expression detected by wildcard expression detection part 203 is “something” unless “su-something” is registered as a wildcard expression in wildcard expression storage part 204, so that the ‘su’ uttered by the user intended as a wildcard expression is treated as if it was a part of a non-wildcard expression. In such case, too, whether or not a part handled as a part of a wildcard is present within a non-wildcard expression is determined in order to allow it to be processed as a part of the wildcard expression when said part does exist.

[0089]

First, a buffer for storing a Mauler code string is prepared for each wildcard expression part detected. When information is detected as a non-wildcard expression part and a wildcard expression part, said buffer is used to store said information in the form of Mauler codes. In addition, because the information detected may appear either before or after the wildcard expression part, 2 buffers are prepared for each wildcard expression to handle said situation. In Step S205, Mauler codes to be input to the buffers are detected for the respective parts of the non-wildcard expression parts adjoining the wildcard expression part. Figure 15 shows an outline of the processing (processing in Step S205 applied to one of the non-wildcard expression parts detected). The explanation will continue with reference to Figure 15 hereafter.

[0090]

Step S205-1

In this step, how the non-wildcard expressions are pronounced is checked. For example, speech analysis part 201 is requested to carry out speech recognition of a target non-wildcard expression part by the unit of a syllable. A Mauler code string output in this step will be referred to as a “partial syllable recognition result” in the explanation given below.

---

\* [Editor's note: "Su" is the first syllable of the Japanese pronunciation of "Stayin."]

[0091]

Step S205-2

In this step, whether there is a partial recognition result which corresponds to the current target non-wildcard expression or not is checked. If no partial recognition result is found to exist as a result, [it advances] to Step S205-3, or to Step 205-4 if a partial recognition result is present.

[0092]

Step S205-3

If there is no partial recognition result which corresponds to the current target non-wildcard expression, a judgment that said non-wildcard expression is a shorter expression than a vocabulary item in substitution expression storage part 206 can be made. Thus, the entire non-wildcard expression is considered as a part of an adjoining wildcard expression, whether said non-wildcard expression is found at the front of the wildcard expression or at the end is determined, its Mauler codes (string) [are stored] the corresponding buffer, the length equivalent to the number of the Mauler codes stored in the buffer is added to the “character string length of wildcard expression” information received from wildcard expression detection part 203 if the wildcard expression is a rhythmic word, and the “wildcard expression position” information is decremented by 1 if said non-wildcard expression part is present at the front of the wildcard expression part before finishing [the processing].

[0093]

Step S205-4)

Here, whether or not any words uttered other than the corresponding partial recognition result are contained in the current target non-wildcard expression part is checked. For example, the length of the Mauler code string for the partial syllable recognition result corresponding to the current target non-wildcard expression and the length of the Mauler code string for the partial recognition result are compared. As a result, if the Mauler code string for the partial recognition result is longer, or they are equal, advances to Step S205-5, or to Step 205-6 if the Mauler code string for the partial syllable recognition result is longer.

[0094]

Step S205-5

A judgment that the current target non-wildcard expression does not have any more information than the corresponding partial recognition result is made, and [the processing] is ended without inputting anything into the buffer.

[0095]

Step S205-6

Because the Mauler code string for the partial recognition result corresponding to the current target non-wildcard expression part is shorter than the Mauler code string for the partial syllable recognition result at the same part, a judgment that there is a possibility that a part of the wildcard expression may be uttered [of] the current non-wildcard expression in addition to the corresponding partial recognition result is made, and to what part of the non-wildcard expression part the partial recognition result corresponds is checked.

[0096]

For example, as shown in Figure 16, it is obtained by comparing their Mauler code strings by checking the Mauler code string for the partial recognition result against the Mauler code string for the partial syllable recognition result in sequence. In Figure 16, although the partial recognition result “Tokyo” (Mauler code string “to-o-kyo-o”\*) and the partial syllable recognition result “to-o-kyo-o-su” are compared, because the length of the Mauler code string for the partial recognition result is 4, and the length of the Mauler code string for the partial syllable recognition result is 5, 2 patterns in which either “to” or “o” (first o) of “to-o-kyo-o-su” are selected as the code used for the matching are feasible for the partial recognition result. When the partial recognition result is even shorter, a pattern in which that after “kyo” is selected as the beginning is also created.

[0097]

Then, which of the patterns is the best for the matching is decided, and what constitutes the “remaining” part is decided. Said “remaining” part is considered as a part which should be taken as a part of the wildcard expression part contained in the non-wildcard expression. In terms of a method for deciding the remaining part, when the number of matched Mauler codes is used as a reference, the part where the largest number of matching Mauler codes are present is selected as the position where the partial recognition result is present, and the part of the partial recognition result where the Mauler code string does not fit is extracted as “remaining,” for example.

/12

[0098]

In the case of the example in Figure 16, “su” is extracted as the remaining part.

---

\* [Editor's note: This spelling corresponds to the Japanese pronunciation of "Tokyo."]

[0099]

In addition, when the position of the partial recognition result cannot be decided uniquely due to the presence of 2 or more patterns by which the number of matching Mauler codes becomes the maximum, a judgment that there is no remaining part is made. Alternatively, the judgment that there is no remaining part may be made also when the number of matching Mauler codes is equal to or lower than a threshold.

[0100]

Step S205-7

Here, whether or not a remaining part was able to be extracted as a result of Step S205-6 is checked. [It advances] to Step S205-8 if a remaining part is present, or to Step S205-5 if no remaining part is present.

[0101]

Step S205-8

Here, whether or not the remaining part as a result of Step S205-6 is present at a position adjoining the wildcard expression part is checked. In the case of the example in Figure 16, because the remaining part “su” is present immediately before the wildcard expression “something,” a judgment that the remaining part “su” is present at a position adjoining “something” is made. To the contrary, when “something” is present before “to-o-kyo-o-su,” a judgment that the remaining “su” found at the end of “to-o-kyo-o-su” is not adjoining “something” is made. In this case, if the “to” is “to-o,” a judgment that a remaining part is present immediately after the wildcard expression “something” can be made. If a remaining is present at the adjoining part, it advances to Step S205-9. If no remaining is present at the adjoining part, it advances to Step S205-5.

[0102]

Step S205-9

Here, because the remaining part extracted in Step S205-6 is adjoining the wildcard expression, said extracted remaining part is considered as a part of the adjoining wildcard expression, whether said remaining part is present at the beginning or the end of the adjoining wildcard expression is determined, its Mauler codes (string) are stored into the corresponding buffer, and [the length] equivalent to the number of the Mauler codes stored in the buffer is added to the “character string length of wildcard expression” information received from wildcard expression detection part 203 if the wildcard expression part is a rhythmic word before ending [the processing].



[0103]

As an alternative to the aforementioned method, speech analysis part 201 may be provided with the word detection function of wildcard expression detection part 203; whereby, Mauler codes (string) to become a part of the wildcard expression in the aforementioned manner can be estimated by detecting the word in the speech recognition result obtained in Step S204 out of [the part] separated from the original signal, separating the signal remaining at the boundary part with the wildcard expression subsequently, and requesting speech analysis part 201 to carry out the speech recognition by the unit of a syllable.

[0104]

Step S206

The word corresponding to the wildcard expression part is retrieved from vocabulary items stored in substitution expression storage part 206 so as to match the vocabulary item stored in vocabulary storage part 202 using the pieces of information obtained through the processing in Steps S204 and S205 and the information shown in Figure 13 obtained from wildcard expression detection part 203 as retrieval conditions. Figure 17 is a flowchart for the operations to be carried out in Step S206. The processing flow will be explained below with reference to Figure 17.

[0105]

Step S206-1

In this step, vocabulary items to be output which match the speech recognition result received are selected from vocabulary storage part 202. For example, words which match the positional conditions of partial recognition results of non-wildcard expression parts are selected utilizing the representation information of the information (Figure 12) stored in vocabulary storage part 202 while using non-wildcard expression parts as conditions and assuming that a wildcard expression is longer than 1 word. Then, a part replaced by the wildcard expression comprising the vocabulary item selected is retrieved from the expression stored in substitution expression storage part 206. For example, assuming that the partial recognition results obtained are "Tokyo" and Hotel," and the sequence of "Tokyo (wildcard expression) Hotel" is found based on the arrangement of the separated [parts of] the original signal and the information on the position of the wildcard expression, "Tokyo Stayin Hotel," for example, in which the word "Tokyo" is present at the beginning, the word "Hotel" is present at the end, and at least 1 word is present between them, is selected as a matching vocabulary item. Then, the expression "Stayin," for example, is retrieved for the part replaced by the wildcard expression from substitution expression storage part 206.

[0106]

Step S206-2

In this step, the expressions extracted in Step S206-1 are further restricted using the Mauler codes (string) stored in the buffer in Step S205 as a retrieval condition.

[0107]

Step S206-3

In this step, whether the wildcard expression contained in the speech recognition result is a multiword substitution word or a rhythmic word is determined. This can be checked using the “wildcard expression type” information in Figure 13 received from wildcard expression detection part 203. Then, if it is a rhythmic word, it advances to Step S206-4. If it is a multiword substitution word, the expression extracted in Step S206-2 and the normal vocabulary item comprising the partial recognition result are output, and the processing is ended. Furthermore, if multiple vocabulary items to be output are present, some or all of them may be output, and processing for presenting multiple answers to let a user make a selection is decided according to processing (200 in the figure) specific to the application at the output destination.

/13

[0108]

Step S206-4

In this step, the Mauler code string length information of the wildcard expression given by wildcard expression detection part 203 and the Mauler code length information stored in substitution expression storage part 206 are compared with respect to the vocabulary items extracted in Step S206-1 in order to further restrict the vocabulary items output. For example, only the items which indicates a Mauler code string length difference less than a given threshold are extracted. Furthermore, if the vocabulary items are not restricted during said processing, the vocabulary items may be prioritized for output.

[0109]

Step S206-5

In this step, the “metrical parameter ” information received from speech analysis part 101 and the “metrical parameter ” information stored in metrical information storage part 104 are compared with respect to the vocabulary items extracted in Step S206-4 in order to decide a vocabulary item to be output. For example, DP method based matching is carried out using the method disclosed in “Keyword Spotting Utilizing Pitch Pattern Information” (collection of lectures by the Acoustical Society of Japan, September Heisei 8[1996], pp. 29-30). Furthermore, although said comparison method differs depending on the metrical parameter configured, an

arbitrary metrical comparison method may be utilized in the present embodiment as long as the configured parameter can be utilized. Then, a vocabulary item with the metrical information most closely resembling that of the speech uttered is output before ending the processing. Alternatively, if there are multiple candidates, they may be prioritized for output in the order of the resemblance of their metrical information.

[0110]

Above are the configurations and the function of substitution expression verification part 205 and the processing method pertaining to the present invention.

[0111]

Next, the aforementioned speech input interpretation method will be explained in further detail below. Here, the example of the map information system used to explain the first application example will be used to explain a specific application example of its operations when speech is input by a user.

[0112]

Assume that information on 4 hotels (Tokyo Stayin Hotel, Tokyo Marunoguchi hotel, Palace Hotel, and Tokyo Inter Continental Hotel) near Tokyo railroad station is registered in said map information system, and the information shown in Figure 18 regarding said 4 hotels and the necessary parameters for their speech recognition are stored in vocabulary storage part 202. Then, they are registered into substitution expression storage part 206. Words separated from said vocabulary items and expressions as combinations of the words are shown in Figure 19.

[0113]

In addition, assume that the multiword substitution word “something” is stored in wildcard expression storage part 204 as a wildcard expression.

[0114]

Next, assume that a user wants to make an inquiry regarding “Tokyo Stayin Hotel,” but he/she does not remember the part “Stayin” clearly, and enters the speech input “Tokyo su-something Hotel.”

[0115]

In order to make the notation clear, a waveform signal before a speech recognition result is obtained will be expressed by (...) as in (signal), and a character string after a speech recognition result is obtained will be expressed by (...) as in (character string).

[0116]

Upon receiving said input, a wildcard expression is detected at wildcard expression detection part 203 first. The signal (Tokyo su-something Hotel) contains the wildcard expression (something), and it is detected as a wildcard expression. Then, the information of the kind shown in Figure 20 is given to substitution expression verification part 205.

[0117]

Processing to be carried out by substitution expression verification part 205 is as follows.

[0118]

Step S201

The point that a wildcard expression is present is confirmed based on the presence/absence of wildcard expression information.

[0119]

Step S204

(to-o-kyo-o-su) and (ho-te-ru<sup>\*</sup>) are found to be the necessary parts to be recognized based on the separated [parts of] original signal and the wildcard expression position information. Then, speech analysis part 201 is requested to carry out word recognition with respect to said 2 [parts of the] signal using the vocabulary items stored in substitution expression storage part 206. Assume that "Tokyo" is obtained as a partial recognition result as a result of the recognition of (to-o-kyo-o-su), and "Hotel" as a result of the recognition of (ho-te-ru).

[0120]

Step S205

The [partial] signal (to-o-kyo-o-su) is processed first.

---

<sup>\*</sup> [Editor's note: This spelling corresponds the Japanese pronunciation of "hotel."]

[0121]

Step S205-1

Speech analysis part 201 is requested to recognize the [partial] signal 1(to-o-kyo-o-su) by the unit of a syllable. Assume that the Mauler code string “to-o-kyo-o-su” is obtained as a result.

[0122]

Step S205-2

Because “Tokyo” has been obtained as a result of the recognition of the [partial] signal (to-o-kyo-o-su), it advances to Step S205-4.

[0123]

Step S205-4

The Mauler code string length of the Mauler code string “to-o-kyo-o-su” is 5. Also, assume that the partial recognition result “Tokyo” is stored into substitution expression storage part 206 in the manner shown in Figure 21.

[0124]

Because the partial syllable recognition result “to-o-kyo-o-su” obtained from the signal input is longer than said Mauler code string length, it advances to Step S205-6.

[0125]

Step S205-6

Figure 16 emerges when the syllable recognition result “to-o-kyo-o-su” and the Mauler code string “to-o-kyo-o” of the partial recognition result “Tokyo,” and the Mauler code “su” is detected as a remaining part.

[0126]

Step S205-7

Because the Mauler code “su” is detected as a remaining part, it advances to Step S205-8.

[0127]

Step S205-8

Because the remaining Mauler code “su” is located at the end of the syllable recognition result “to-o-kyo-o-su,” and the original signal (to-o-kyo-o-su) of said syllable recognition result is located immediately before the wildcard expression part (something), the remaining part “su” is judged as a part of the wildcard expression. It advances to Step S205-9.

[0128]

Step S205-9

The Mauler code “su” is input to the buffer which stores the pronunciation of the beginning part of the wildcard expression.

[0129]

Next, the same processing is carried out with respect to the [partial] signal (ho-te-ru). Here, assume that no remaining part is found in the partial recognition result “Hotel,” and it advances to the next step without storing anything into the buffer.

[0130]

Step S206

Here, appropriate vocabulary items are retrieved based on the information obtained thus far.

[0131]

Step S206-1

“Tokyo (wildcard expression) Hotel” is judged to be the target vocabulary item of the speech input based on the original signal information, partial recognition results, and wildcard expression position information. “Tokyo Stayin Hotel,” “Tokyo Marunoguchi Hotel,” and “Tokyo Inter Continental Hotel” are selected from the vocabulary items stored in vocabulary storage part 202 as vocabulary items which match the aforementioned condition. In addition, “Stayin,” “Marunoguchi,” and “Inter Continental” are selected among the expressions stored in substitution expression storage part 206 as candidates to be output as expressions replaced by the wildcard expression. At this time, “Palace Hotel” is either excluded as a candidate to be output or considered as a candidate with a lower priority.

[0132]

Step S206-2

A judgment that the expression “Stayin” which begins with the Mauler code “su” is promising can be made in reference to [the content] stored in the buffer in Step S205. Here, the vocabulary item “Tokyo Stayin Hotel” containing “Stayin” becomes promising as a candidate to be output. “Tokyo Marunoguchi Hotel” and “Tokyo Inter Continental Hotel” are either excluded as candidates to be output or considered as candidates with a lower priority.

[0133]

Step S206-3

Because it is clear that the wildcard expression (“something”) used is a multiword substitution word based on the information sent from wildcard expression detection part 203, “Tokyo Stayin Hotel” is output as the first candidate. Alternatively, when the application part corresponds to multiple candidates, “Tokyo Marunoguchi Hotel” and “Tokyo Inter Continental Hotel” are output, and “Palace Hotel” may also be output as a candidate with even an lower priority. Then, upon receiving said output, the application-specific processing (200 in the figure) carries out appropriate processing.

[0134]

Now, the processing of the speech input “Tokyosu-something-hotel” is completed.

[0135]

From the explanation given above, it is clear that in the case of the speech input analyzer pertaining to the present embodiment, when a user enters the speech input “Tokyo something Hotel” by uttering the specific parts he/she remembers while using a wildcard expression for the uncertain part, even though the user does not remember the name “Tokyo Stayin Hotel” clearly, it can be interpreted into an appropriate name so as to output the information to the application part, and that the speech information input by the user is utilized effectively in that when “Tokyo something Hotel” is input while using a wildcard expression for the unclear part in “Tokyo su ... Hotel” on which the user lacks detailed information, “Tokyo Stayin Hotel” is given a higher priority than “Tokyo Marunoguchi Hotel” and “Tokyo Inter Continental Hotel” with the same format, that is, “Tokyo ... Hotel.”

[0136]

As such, with the present device configured in the aforementioned manner, a speech input interpreter can be constructed which is capable of operating even when a user does not remember correct words or sentences he/she is allowed to utter.

[0137]

For example, a speech input interpreter can be constructed by which a recognition error can be restrained, and the output by the system for speech input can be directed toward the intention of the user even when the user remembers only a portion of acceptable words or a sentence.

[0138]

In addition, a speech input interpreter can be constructed by which a recognition error can be restrained, and an output of the system for speech input can be directed to the intention of the user even when the user remembers only the “rhythm” of acceptable words or a sentence.

[0139]

Furthermore, the functions and the effects of the respective embodiments are not restricted to the aforementioned examples. For example, an operation error can be avoided by presenting a list of the results of the substitution processing at substitution expression verification part 103 in the first embodiment, or at substitution expression verification part 205 in the second embodiment, in order for the user to make a correct choice.

[0140]

In addition, it may be utilized as an input means for a multimodal interface while narrowing down the search range and reducing the redundancy of the output in order to reduce the burden on the user.

[0141]

In addition, it may be utilized as an input means not only for a multimodal interface but also for a device involving an arbitrary speech input. In addition, the metrical information may be used for the analysis of all speech information input without restricted to the part where a wildcard expression is used.

[0142]

A device configuration for realizing the processing by the present speech input interpreter using software will be explained below with reference to Figure 22.

[0143]

In this case, the hardware part of the present speech input interpreter comprises CPU 21, RAM 22 for storing necessary programs and data, disk drive device 24, memory device 25, and input input/output device 26.

[0144]

In the case of the first embodiment, speech analysis part 101, vocabulary storage part 102, substitution expression verification part 103, and metrical information storage part 104 in Figure 1 are configured using programs containing the descriptions of the respective processing steps.



[0145]

In the case of the second embodiment, speech analysis part 201, vocabulary storage part 202, wildcard expression detection part 203, wildcard expression storage part 204, substitution expression verification part 205, and substitution expression storage part 206 in Figure 11 are configured using programs containing the descriptions of the respective processing steps.

/15

[0146]

Furthermore, the information to be stored in each buffer may be formed either integrally with a program or separately from the program.

[0147]

Said programs containing the descriptions of the respective processing steps are stored in RAM 22 as programs for controlling the computer system shown in Figure 22 and executed by CPU 21. CPU 21 carries out arithmetic operations and the control over memory device 25 or input input/output device 26 according to the program procedures stored in RAM 22 in order to realize desired functions.

[0148]

A variety of methods can be used to install the programs into RAM 22. For example, the aforementioned programs (programs containing the descriptions of the processing steps carried out by speech analysis part 101, vocabulary storage part 102, substitution expression verification part 103, and metrical information storage part 104 in Figure 1 and used to control the computer system and programs containing the descriptions of the processing steps carried out by speech analysis part 201, vocabulary storage part 202, wildcard expression detection part 203, wildcard expression storage part 204, substitution expression verification part 205, and substitution expression storage part 206 in Figure 11 and used to control the computer system) is stored in advance on a storage medium (for example, a floppy disk or a removable recording medium such as a CD-ROM) which can be read by the computer. Then, as shown in Figure 22, said programs are read using a disk drive device appropriate for the recording medium and stored into RAM 22. Alternatively, it is installed onto disk drive device 24 once and stored into RAM 22 from said device when it is to be executed.

[0149]

In addition, when the recording medium used to store the programs is an IC card, said programs can be read using an IC card reader. Furthermore, the programs can be received through a prescribed interface device via a network also.

[0150]

Furthermore, an application which utilizes the interpretation results may be installed on the speech input interpreter, or the speech input interpreter and the device used to install the application may be made independent of each other. In addition, the programs for realizing the speech input interpreter and the program for realizing the application which utilizes the interpretation results may be executed by the same CPU, or they may be executed using CPUs provided separately.

[0151]

Incidentally, although the first and the second embodiments are described as if they were based on the assumption that only 1 wildcard expression would be input, even when multiple wildcard expressions are input, they can be handled by generating the corresponding vocabulary items at vocabulary storage part 102 and applying the same processing to the respective applicable wildcard expression parts at substitution expression verification part 103 in the case of the first embodiment, and they can be handled by applying the same processing (that is, information on the positions, types, and rhythms of the detected multiple wildcard expressions is given to substitution expression verification part 205, one for storing [a part] within a wildcard expression is added to the buffer for storing a part of a wildcard expression, wildcard expressions which appear continuously are handled as 1 wildcard expression) to the respective wildcard expressions detected in the case of the second embodiment.

[0152]

In addition, the retrieval conditions set in the first and the second embodiments are not unique to the respective embodiments. For example, the speech input time may be utilized at the time of the retrieval of a substitution expression in the second embodiment. In addition, although the first embodiment is described as if it was based on the assumption that an input containing a correct expression as a part of a wildcard expression, for example, "su'-something," would never be input, it can be handled easily by setting a vocabulary item, such as "su/something," at vocabulary storage part 102. In addition, rhythms of all expressions can be used as retrieval conditions instead of defining a wildcard expression as a multiword substitution word or as a rhythmic word.

[0153]

In addition, the present invention can be applied not only to the Japanese language but also to all languages which contain wildcard expressions using a common unit for analysis, such as a syllable or a phoneme, during the analysis by the unit of a Mauler code. In addition, the present invention can be used also to retrieve [musical notes] of a song sung using rhythms for the parts where the musical notes are unknown, for example.

[0154]

The present invention is not restricted to the aforementioned embodiments, and it can be implemented in a variety of ways without exceeding its technical scope.

[0155]

#### Effect of the invention

In the present invention, because a part where a part of a normal vocabulary item is replaced is detected from an input speech, and it is substituted with a normal expression appropriate for said part, even when a user does not remember acceptable vocabulary items clearly, an speech input containing said replacement expression can be accepted and interpreted.

#### Brief description of the figures

Figure 1 is a diagram showing an example of the configuration of the speech input interpreter pertaining to a first embodiment of the present invention.

Figure 2 is a diagram showing an example of the information stored in the vocabulary storage part.

Figure 3 is a diagram showing an example of the vocabulary items stored in the vocabulary storage part.

Figure 4 is a diagram showing an example of the information given from the speech analysis part to the substitution expression verification part.

Figure 5 is a diagram showing an example of the information stored in the metrical information storage part.

Figure 6 is a flowchart showing an example of the operations carried out by the substitution expression verification part.

Figure 7 is a diagram showing an example of the vocabulary items registered to the vocabulary storage part.

Figure 8 is a diagram showing an example of the information registered to the metrical information storage part.

Figure 9 a diagram showing an example of the information output from the speech analysis part to the substitution expression verification part.

Figure 10 is a diagram showing an example of the result of the retrieval of vocabulary items which match the speech recognition result.

/16

Figure 11 is a diagram showing an example of the configuration of the speech input interpreter pertaining to a second embodiment of the present embodiment.

Figure 12 is a diagram showing an example of the information stored in the vocabulary storage part.

Figure 13 is a diagram showing an example of the information given from the wildcard expression detection part to the substitution expression verification part.

Figure 14 is a flowchart showing an example of the operations carried out by the substitution expression verification part.

Figure 15 is a flowchart showing an example of the processing procedure applied to a non-wildcard expression part.

Figure 16 is a diagram for explaining the retrieval of a part of a wildcard expression.

Figure 17 is a flowchart showing an example of the processing procedure applied to a wildcard expression part.

Figure 18 is a diagram showing an example of the vocabulary item registered in the vocabulary storage part.

Figure 19 is a diagram showing an example of the information registered in the substitution expression verification part.

Figure 20 is a diagram showing an example of the information given from the wildcard expression detection part to the substitution expression verification part.

Figure 21 is a diagram showing an example of the information stored in the substitution expression verification part.

Figure 22 is a diagram showing an example of the hardware configuration.

#### Explanation of reference numbers

- 1, 2. speech input interpreter
- 100. speech input device
- 101. speech analysis part
- 102. vocabulary storage part
- 103. substitution expression verification part
- 104. metrical information storage part
- 201. speech analysis part
- 202. vocabulary storage part

- 203. wildcard expression detection part
- 204. wildcard expression storage part
- 205. substitution expression verification part
- 206. substitution expression storage part
- 21. CPU
- 22. RAM
- 23. bus
- 24. disk drive device
- 25. memory device
- 26. input/output device

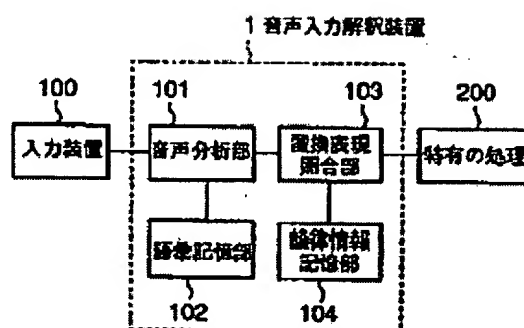


Figure 1

- Key:
- 1 Speech input interpreter
  - 100 Speech input device
  - 101 Speech analysis part
  - 102 Vocabulary storage part
  - 103 Substitution expression verification part
  - 104 Metrical information storage part
  - 200 Specific processing

| Information                             | Example               |
|---|-----------------------|
| Representation                          | Tokyo/honyalala/Hotel |
| Presence/absence of wildcard expression | Yes                   |
| Expression type                         | (OK/replace/OK)       |
| Wildcard expression type                | Rhythmic word         |
| Speech recognition parameter            |                       |

Figure 2

|                         |                           |
|-------------------------|---------------------------|
| Tokyo/Stayin/Hotel      | Something/Stayin/Hotel    |
| Tokyo/something/Hotel   | Tokyo/Stayin/something    |
| Something/Hotel         | Tokyo/something           |
| Honyalala/Stayin/Hotel  | Tokyo/honyalalala/Hotel   |
| Tokyo/Stayin/honyala    | Honyalalalalalalala/Hotel |
| Tokyo/honyalalalalalala |                           |

Figure 3

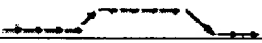
| Information                             | Example  |
|---|--|
| Recognition result                      | “Tokyo/honyalalala/Hotel”  |
| Word utterance time                     | (650 msec/820 msec/510 msec)   |
| Metrical parameter                      |  |
| Presence/absence of wildcard expression | Yes  |
| Expression type                         | (OK/replace/OK)  |
| Wildcard expression type                | Rhythmic word  |

Figure 4

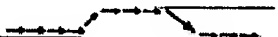
| Information    | Example  |
|----------------|--|
| Representation | “Tokyo/Stayin/Hotel”   |
| Standard time  | (700 msec/900 msec/550 msec)   |
| Rhythm         |  |

Figure 5

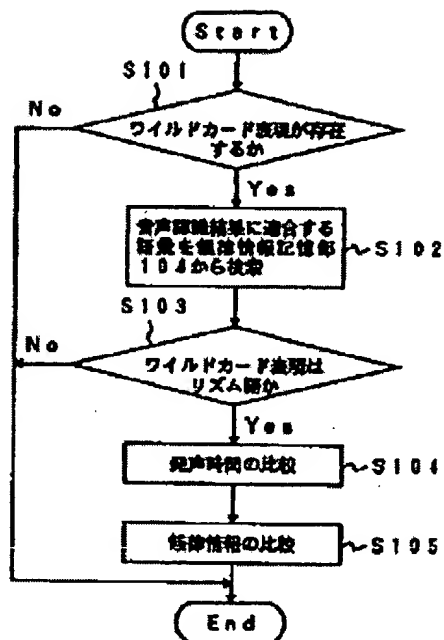


Figure 6

- Key: S101 Is a wildcard expression present?  
 S102 Retrieve vocabulary items which match the speech recognition result from metrical information storage part 104.  
 S103 Is the wildcard expression a rhythmic word?  
 S104 Compare its utterance time.  
 S105 Compare its metrical information.

|   |   |
|---|---|
| Palace/Hotel                              | Tokyo/Stayin/Hotel                        |
| Tokyo/Marunoguchi/Hotel                   | Tokyo/Inter/Continental/Hotel             |
| Honyala/Hotel                             | Palace/honyala                            |
| Honyalala/Stayin/Hotel                    | Tokyo/honyalalala/Hotel                   |
| Tokyo/Stayin/honyala                      | Honyalala/Marunoguchi/Hotel               |
| Tokyo/Marunoguchi/honyala                 | Honyalalalalalalala/Hotel                 |
| Tokyo/honyalalalalalalala                 | Honyalala/Inter/Continental//Hotel        |
| Tokyo/honyalala/Continental/Hotel         | Tokyo/Inter/honyalalalalalala/Hotel       |
| Tokyo/Inter/Continental/honyala           | Honyalalalalalalala/Continental/Hotel     |
| Tokyo/honyalalalalalalalalalala/Hotel     | Tokyo/Inter/honyalalalalalalalalalala     |
| Honyalalalalalalalalalalalalalalala/Hotel | Tokyo/honyalalalalalalalalalalalalalalala |

Figure 7


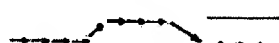
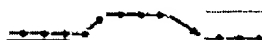

|                      |   |   |
|----------------------|---|---|
| Vocabulary item      | Palace Hotel  | Tokyo Stayin Hotel  |
| Representation       | Palace/Hotel  | Tokyo/Stayin/Hotel  |
| Standard time        | (450 msec/ 550 msec)  | (700 msec/ 900 msec /550 msec)  |
| Metrical information |  |  |
| Vocabulary item      | Tokyo Marunoguchi Hotel   | Tokyo Inter Continental Hotel   |
| Representation       | Tokyo/Marunoguchi/Hotel   | Tokyo/Inter/Continental/Hotel   |
| Standard time        |   |   |
| Metrical information |  |  |

Figure 8

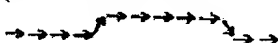
|   |  |
|---|--|
| Recognition result                      | Tokyo/honyalalala/Hotel  |
| Word utterance time                     | (650 msec/ 820 msec /510 msec)   |
| Metrical parameter                      |  |
| Presence/absence of wildcard expression | Yes  |
| Expression type                         | (OK/replace/OK)  |
| Wildcard expression type                | Rhythmic word  |

Figure 9

|                  | ② 比 率                      |                            | 伸長係数   | 伸長後<br>(msec) | 標準時間<br>(msec) |
|------------------|----------------------------|----------------------------|--------|---------------|----------------|
|                  | ③「東京」                      | ④「ホテル」                     |        |               |                |
| ① 東京ステインホテル      | $\frac{700}{650} = 1.0769$ | $\frac{550}{510} = 1.0784$ | 1.0777 | 884           | 900            |
| 東京丸の口ホテル         | $\frac{680}{650} = 1.0615$ | $\frac{580}{510} = 1.0980$ | 1.0798 | 885           | 850            |
| 東京エンターコンチネンタルホテル | $\frac{700}{650} = 1.0769$ | $\frac{560}{510} = 1.0980$ | 1.0875 | 892           | 1700           |

Figure 10

- Key: 1 Tokyo Stayin Hotel  
Tokyo Marunoguchi Hotel  
Tokyo Inter Continental Hotel  
2 Ratio  
3 Tokyo  
4 Hotel  
5 Expansion factor  
After expansion  
Standard time



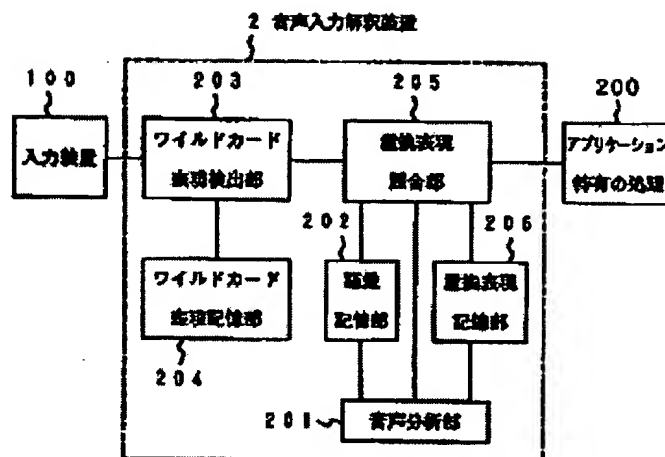


Figure 11

|      |     |   |
|------|-----|---|
| Key: | 2   | speech input interpreter                  |
|      | 100 | speech input device                       |
|      | 200 | application-specific processing           |
|      | 201 | speech analysis part                      |
|      | 202 | vocabulary storage part                   |
|      | 203 | wildcard expression detection part        |
|      | 204 | wildcard expression storage part          |
|      | 205 | substitution expression verification part |
|      | 206 | substitution expression storage part      |

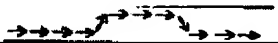
| Information                                | Example  |
|--|--|
| Representation character string            | Tokyo Stayin Hotel   |
| Mauler code string                         | To-o-kyo-o-su-te-e-i-n-ho-te-ru*   |
| Mauler code length                         | 12   |
| Metrical parameter                         |  |
| Necessary parameter for speech recognition |  |

Figure 12

\* [Editor's note: This spelling corresponds to the Japanese pronunciation of Tokyo Stayin Hotel.]

| Information                                      | Example                     |
|--|-----------------------------|
| Presence/absence of wildcard expression          | Yes                         |
| Original signal                                  | "Tokyo" "honyalala" "Hotel" |
| Position of wildcard expression                  | 2                           |
| Wildcard expression type                         | Rhythmic word               |
| Mauler code string length of wildcard expression | 4                           |
| Metrical information on wildcard expression      | ↗→→→                        |

Figure 13

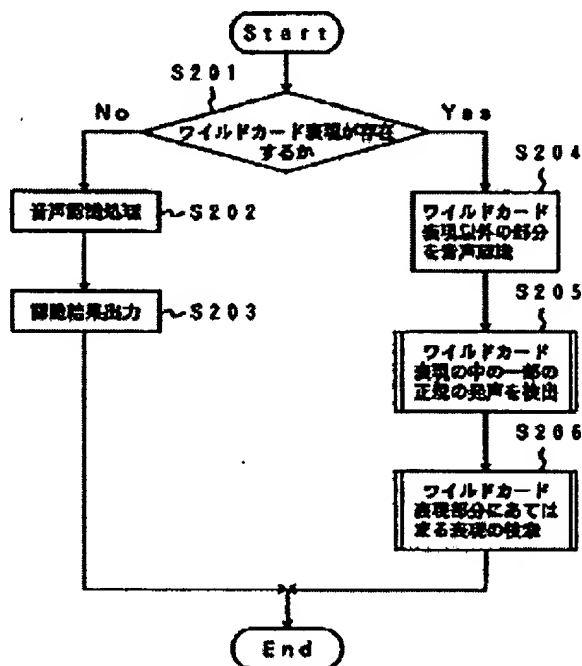


Figure 14

- Key:
- S201 Is a wildcard expression present?
  - S202 speech recognition processing
  - S203 Output recognition result.
  - S204 Apply speech recognition to the part other than the wildcard expression.
  - S205 Detect a partial normal utterance within the wildcard expression.
  - S206 Retrieve an expression corresponding to the wildcard expression part.

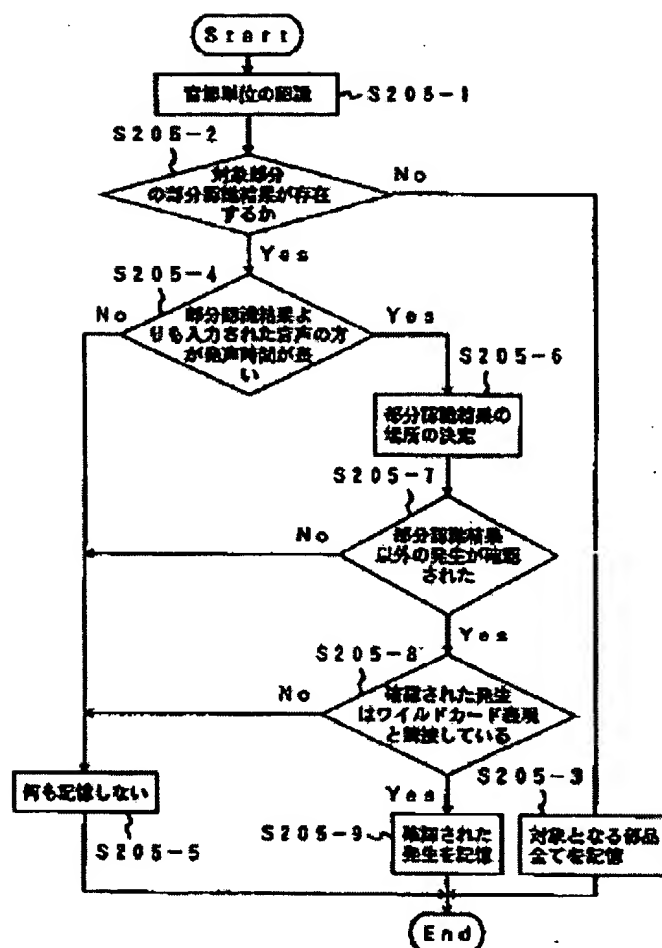


Figure 15

|      |        |  |
|------|--------|--|
| Key: | S205-1 | recognition by the unit of a syllable                          |
|      | S205-2 | Is a partial recognition result of the target part present?    |
|      | S205-3 | Store all target parts.  |
|      | S205-4 | Input speech is longer than the partial recognition result.    |
|      | S205-5 | Nothing is recorded.   |
|      | S205-6 | Decide the position of the partial recognition result.         |
|      | S205-7 | Occurrence other than the partial recognition result is found. |
|      | S205-8 | The occurrence found is adjoining wildcard expression.         |
|      | S205-9 | Store the occurrence found.                                    |

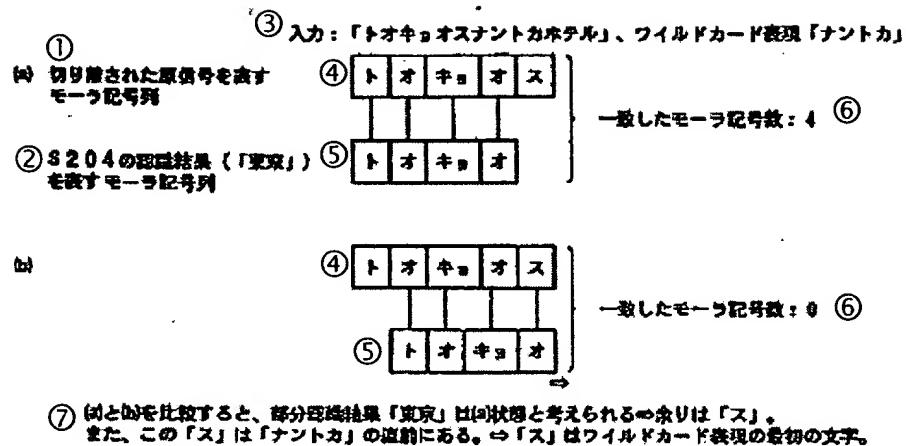


Figure 16

- Key:
- 1 Mauler code string representing separated [parts of] the original signal
  - 2 Mauler code string representing recognition result ("Tokyo") of S204
  - 3 Input: "Tokyo su-something-Hotel", Wildcard expression "something"
  - 4 to-o-kyo-o-su
  - 5 to-o-kyo-o
  - 6 number of Mauler codes matched
  - 7 When (a) and (b) are compared, partial recognition result "Tokyo" is considered to be under condition (a). ⇒ Remaining part is "su."
- Also, this "su" is placed immediately before "something." ⇒ "su" is the first character of the wildcard expression.

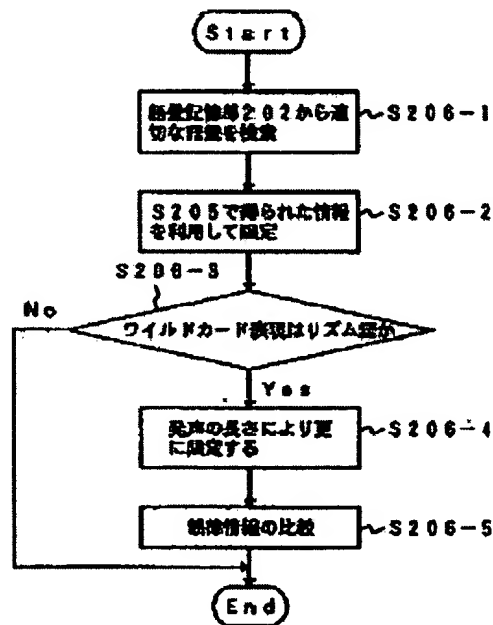


Figure 17

- Key: S206-1 Retrieve appropriate vocabulary items from vocabulary storage part 202.  
 S206-2 Restrict them using the information obtained in S205.  
 S206-3 Is the wildcard expression a rhythmic word?  
 S206-4 Further restrict them based on the utterance length.  
 S206-5 Compare its metrical information.

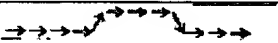
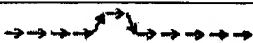
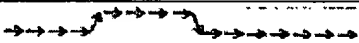
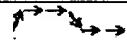
|                                 |   |   |
|---------------------------------|---|---|
|                                 | Tokyo Stayin Hotel  | Tokyo Marunoguchi Hotel   |
| Representation character string | Tokyo Stayin Hotel  | Tokyo Marunoguchi Hotel   |
| Mauler code string              | To-o-kyo-o-su-te-e-i-n-ho-te-ru   | To-o-kyo-o-ma-ru-no-gu-chi-ho-te-ru   |
| Mauler code length              | 12  | 12  |
| Metrical parameter              |    |    |
|                                 | Tokyo Inter Continental Hotel   | Palace Hotel  |
| Representation character string | Tokyo Inter Continental Hotel   | Palace Hotel  |
| Mauler code string              | To-o-kyo-o-e-n-ta-a-ko-n-chi-ne-n-ta-ru-ho-te-ru*                                   | Palace Hotel  |
| Mauler code length              | 18  | 6   |
| Metrical parameter              |  |  |

Figure 18

|                         |                         |                   |
|-------------------------|-------------------------|-------------------|
| Tokyo                   | Tokyo Stayin            | Tokyo Marunoguchi |
| Continental             | Continental Hotel       | Palace            |
| Stayin                  | Stayin Hotel            | Marunoguchi Hotel |
| Tokyo Inter             | Tokyo Inter Continental | Hotel             |
| Marunoguchi             | Inter                   | Inter Continental |
| Inter Continental Hotel |                         |                   |

Figure 19

\* [Editor's note: This spelling corresponds to the Japanese pronunciation of "Tokyo Intercontinental Hotel."]

| Information type                               | Content                                |
|--|--|
| Presence/absence of wildcard expression        | Yes                                    |
| Original signal                                | (to-o-kyo-o-su) (something) (ho-te-ru) |
| Position of wildcard expression                | 2                                      |
| Wildcard expression type                       | Multiword substitution word            |
| Character string length of wildcard expression | -                                      |
| Metrical information on wildcard expression    | -                                      |

Figure 20

|                    |            |
|--------------------|------------|
| Representation     | Tokyo      |
| Mauler code string | To-o-kyo-o |
| Mauler code length | 4          |
| Metrical parameter | →→→→       |

Figure 21

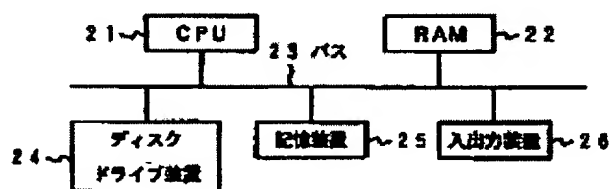


Figure 22

Key: 23 bus  
 24 disk drive device  
 25 memory device  
 26 input/output device

(Translation)

**OFFICIAL NOTICE OF REJECTION**

Patent Application No. 2000-380781

Examiner, The Patent Office: T. Kashimoto(9379 5C00)

Date of Draft: December 25, 2002

Mailing Date: January 7, 2003

Applicant: SECRETARY OF AGENCY OF INDUSTRIAL SCIENCE AND TECHNOLOGY

Applied Provision(s): Section 29 Paragraph 2

The present application is rejected for the following reasons. The applicant may present an argument, if any, within 60 days from the mailing date of this Official Notice.

**REASONS**

The invention(s) of the present application as claimed in claim(s) set forth below would have been obvious to one skilled in the art, to which the invention(s) pertain(s), prior to the filing date of the present application, on the basis of the publication(s) set forth below as distributed in Japan or foreign countries prior to the filing date of the present application. Thus, the present invention(s) is(are) unpatentable under the provision of Section 29, Paragraph 2 of the Patent Law.

Remarks (refer to a reference list below)

1. Claims 1 to 14 and References 1 and 2

Reference 1 discloses that a specific keyword phonated before or after a character string is regarded as an arbitrary character string, and then, the keyword portion is complemented. Reference 1 also discloses outputting a list of complement candidates if there are a plurality of complement candidates. Further, reference 2 discloses that a portion subsequent to a fragment of a word is complemented when triggered by a user's intentional "vocal pausing after phonating a fragment of a word".

**Reference List**

1. Japanese Patent Application Laid-open No. Hei 11-95793
2. Japanese Patent Application Laid-open No. Hei 10-222337  
(In particular, refer to claim 7 and paragraphs [0093] to [0099].)

Record of Search Result of Prior Art Literature

Searched Field

IPC seventh edition      G10L15/00 - 15/28

Prior Art Literature

Goto, Ito, Hayamizu, A Synonym for "A Real-time Detection System for a Hesitation Pause in Natural Conversation", Research Report of Information Processing Society of Japan (Speech Information Processing), 27-2, pp. 9-16, July 23, 1999)



整理番号 10744211

発送番号 437455

発送日 平成15年 1月 7日 1 / 2

## 拒絶理由通知書



特許出願の番号

特願2000-380781

起案日

平成14年12月25日

特許庁審査官

檜本 剛

9379 5C00

特許出願人

独立行政法人産業技術総合研究所 (外 2名)

様

適用条文

第29条第2項

この出願は、次の理由によって拒絶をすべきものである。これについて意見があれば、この通知書の発送の日から60日以内に意見書を提出して下さい。

## 理 由

この出願の下記の請求項に係る発明は、その出願前日本国内又は外国において頒布された下記の特許公報に記載された発明に基いて、その出願前にその発明の属する技術の分野における通常の知識を有する者が容易に発明をすることができたものであるから、特許法第29条第2項の規定により特許を受けることができない。

記 (引用文献等については引用文献等一覧参照)

1. 請求項1～14 / 引用文献1～2

ある文字列の前又は後に発声された特定キーワードを任意の文字列と見なし、該キーワード部分を補完する点、及び、補完候補が複数ある場合に補完候補一覧を出力する点は、引用文献1に、ユーザの意図的な「単語断片発声後の有声休止」をトリガーとして、単語断片より後の部分を補完する点は、引用文献2に、それぞれ記載されている。

## 引 用 文 献 一 覧

1. 特開平11-95793号公報

2. 特開平10-222337号公報

(特に【請求項7】及び【0093】～【0099】を参照)



発送番号 437455

2 / 2

先行技術文献調査結果の記録

調査した分野 IPC第7版 G10L15/00-15/28

先行技術文献 後藤、伊藤、速水、「自然発話中の言い淀み箇所のリアルタイム検出システム」、情報処理学会研究報告（音声情報処理）、27-2、p.9-16、平成11年7月23日、

（この拒絶理由通知の内容に関するお問い合わせ、または面接のご希望がございましたら、下記までご連絡ください。

特許審査第四部映像機器 榎本 剛

TEL:(03)3581-1101 内線3541 FAX:(03)3501-0715 )

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-95793

(43) 公開日 平成11年(1999)4月9日

|                               |       |              |         |
|-------------------------------|-------|--------------|---------|
| (51) Int.Cl. <sup>8</sup>     | 識別記号  | F I          |         |
| G 1 0 L 3/00                  | 5 7 1 | G 1 0 L 3/00 | 5 7 1 H |
|                               | 5 3 1 |              | 5 3 1 P |
|                               | 5 6 1 |              | 5 6 1 G |
| 5/04                          |       | 5/04         | E       |
| 9/00                          | 3 0 1 | 9/00         | 3 0 1 B |
| 審査請求 未請求 請求項の数12 O L (全 19 頁) |       |              |         |

(21) 出願番号 特願平9-252446

(22) 出願日 平成9年(1997)9月17日

(71) 出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72) 発明者 屋野 武秀

兵庫県神戸市東灘区本山南町8丁目6番26

号 株式会社東芝関西研究所内

(72) 発明者 知野 哲朗

兵庫県神戸市東灘区本山南町8丁目6番26

号 株式会社東芝関西研究所内

(72) 発明者 河野 恭之

兵庫県神戸市東灘区本山南町8丁目6番26

号 株式会社東芝関西研究所内

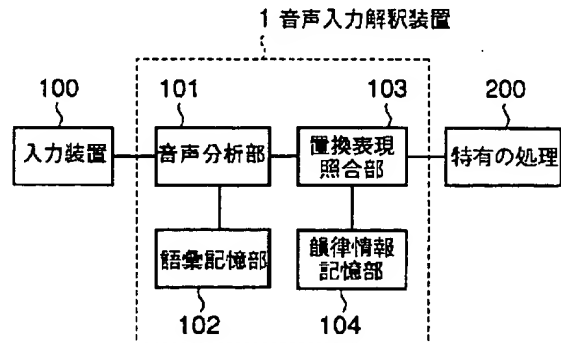
(74) 代理人 弁理士 鈴江 武彦 (外6名)

(54) 【発明の名称】 音声入力解釈装置及び音声入力解釈方法

(57) 【要約】

【課題】 正規の語彙の一部が代替表現で置換された音声入力を解釈して妥当する正規な語彙を求めることの可能な音声入力解釈装置を提供すること。

【解決手段】 入力音声进行解釈して該当する語彙の情報を出力する音声入力解釈装置において、正規の語彙に関する第1の情報、および該正規の語彙の一部が予め定められた代替表現に置き換えられて音声入力されることを考慮した該正規の語彙に関する第2の情報を記憶する手段と、入力音声を音声認識する手段と、前記第2の情報をもとに、前記音声認識結果から前記代替表現を検出する手段と、この手段により前記認識結果から前記代替表現が検出された場合、少なくとも前記入力音声の認識結果に含まれる該代替表現以外の語彙の部分をもとに、前記第1の情報を検索して、該当する語彙を求める手段とを備えたことを特徴とする。



**【特許請求の範囲】**

【請求項1】入力音声を解釈して該当する語彙の情報を出力する音声入力解釈装置において、  
正規の語彙に関する第1の情報、および該正規の語彙の一部が予め定められた代替表現に置き換えられて音声入力されることを考慮した該正規の語彙に関する第2の情報を記憶する手段と、  
入力音声を音声認識する手段と、  
前記第2の情報をもとに、前記音声認識結果から前記代替表現を検出する手段と、  
この手段により前記認識結果から前記代替表現が検出された場合、少なくとも前記入力音声の認識結果に含まれる該代替表現以外の語彙の部分をもとに、前記第1の情報を検索して、該当する語彙を求める手段とを備えたことを特徴とする音声入力解釈装置。

【請求項2】前記該当する語彙が複数検索された場合、少なくとも前記代替表現に対応する音声の音韻的特徴に基づいて、該当する語彙の優先度を評価する手段をさらに備えたことを特徴とする請求項1に記載の音声入力解釈装置。

【請求項3】入力音声を解釈して該当する語彙の情報を出力する音声入力解釈装置において、  
任意の言葉の代替となる代替表現によって音声認識対象となる予め定められた正規の語彙の一部を代替した代替表現を語彙の一種として記憶する語彙記憶手段と、  
前記語彙記憶手段に記憶されている語彙のうち前記代替表現を含まない前記正規の語彙の表記および韻律情報を記憶する韻律情報記憶手段と、  
音声入力装置を介して入力された音声に対し、前記語彙記憶手段を参照して、音声認識および音声の韻律に関する分析を行う音声分析手段と、  
前記音声分析手段による前記入力された音声に対する前記音声認識の結果および前記韻律に関する解析の結果に基づき、前記韻律情報記憶手段を参照して、前記代替表現の部分を前記正規の語彙の部分で置換する置換表現照合手段とを備えたことを特徴とする音声入力解釈装置。

【請求項4】音声入力装置から入力された音声进行分析し、音声認識し、音声認識結果を含む音声分析結果を出力する手段と、該音声認識を行う際に認識対象となる語彙を記憶する語彙記憶手段とを備えた音声入力解釈装置において、  
任意の言葉の代替となる代替表現を記憶する代替表現記憶手段と、  
入力された音声情報から前記代替表現記憶手段に記憶されている語彙と同じ表現を検出する代替表現検出手段と、  
前記語彙記憶手段に記憶されている語彙をさらに分割して別単語としたものを記憶する置換表現記憶手段と、  
前記代替表現検出手段により前記代替表現の検出された入力音声情報における該代替表現でない部分の音声認識

を、前記置換表現記憶手段に記憶されている語彙を音声認識対象として実行し、この音声認識結果を利用して前記置換表現記憶手段に記憶されている語彙から代替表現された言葉として妥当な語彙を検索する処理手段とを備えたことを特徴とする音声入力解釈装置。

【請求項5】前記処理手段は、  
前記音声認識を音節または音韻単位で行い、  
この音節または音韻単位の認識結果を参照することにより、前記代替表現の一部として前記正規の語彙の一部が付加されて発声された部分を検出し、  
前記置換表現記憶手段に記憶されている語彙から代替表現された表現を検索する際に、前記検出結果に適合した表現を優先的に選択することを特徴とする請求項4に記載の音声入力解釈装置。

【請求項6】前記代替表現検出手段は、入力音声の韻律について分析し、  
前記処理手段は、前記置換表現記憶手段に記憶されている語彙から代替表現された表現を検索する際に、前記分析の結果得られた韻律の条件に適合または近似した言葉を優先的に選択することを特徴とする請求項4に記載の音声入力解釈装置。

【請求項7】入力音声を解釈して該当する語彙の情報を出力する音声入力解釈方法において、  
入力音声を音声認識し、  
予め定められた正規の語彙の一部が予め定められた代替表現に置き換えられて音声入力されることを考慮した該正規の語彙に関する情報をもとに、前記音声認識結果から前記代替表現を検出し、  
前記認識結果から前記代替表現が検出された場合、少なくとも前記入力音声の認識結果に含まれる該代替表現以外の語彙の部分をもとに、予め定められた正規の語彙に関する情報を検索して、該当する語彙を求めることを特徴とする音声入力解釈方法。

【請求項8】前記該当する語彙が複数検索された場合、少なくとも前記代替表現に対応する音声の音韻的特徴に基づいて、該当する語彙の優先度を評価することを特徴とする請求項7に記載の音声入力解釈方法。

【請求項9】入力音声を解釈して該当する語彙の情報を出力する音声入力解釈方法において、  
音声入力装置を介して入力された音声に対し、任意の言葉の代替となる代替表現によって音声認識対象となる予め定められた正規の語彙の一部を代替した代替表現を語彙の一種として記憶する語彙記憶手段を参照して、音声認識および音声の韻律に関する分析を行い、  
前記入力された音声に対する前記音声認識の結果および前記韻律に関する解析の結果に基づき、前記語彙記憶手段に記憶されている語彙のうち前記代替表現を含まない前記正規の語彙の表記および韻律情報を記憶する前記韻律情報記憶手段を参照して、前記代替表現の部分を前記正規の語彙の部分で置換することを特徴とする音声入力

解釈方法。

【請求項10】入力音声を音声認識を通じて解釈し、該音声認識を行う際に認識対象となる語彙を記憶する語彙記憶手段のうちの該当する語彙の情報を出力する音声入力解釈方法において、

入力された音声情報から、任意の言葉の代替となる代替表現を記憶する代替表現記憶手段に記憶されている語彙と同じ表現を検出し、

前記代替表現の検出された入力音声情報における該代替表現でない部分の音声認識を、前記語彙記憶手段に記憶されている語彙をさらに分割して別単語としたものを記憶する置換表現記憶手段に記憶されている語彙を音声認識対象として実行し、この音声認識結果を利用して前記置換表現記憶手段に記憶されている語彙から代替表現された言葉として妥当な語彙を検索することを特徴とする音声入力解釈方法。

【請求項11】前記語彙を検索するにあたっては、前記音声認識は音節または音韻単位で行い、この音節または音韻単位の認識結果を参照することにより、前記代替表現の一部として前記正規の語彙の一部が付加されて発声された部分を検出し、前記置換表現記憶手段に記憶されている語彙から代替表現された表現を検索する際に、前記検出結果に適合した表現を優先的に選択することを特徴とする請求項10に記載の音声入力解釈方法。

【請求項12】前記置換表現記憶手段に記憶されている語彙から代替表現された表現を検索する際に、入力音声の韻律について分析を行った結果得られた韻律の条件に適合または近似した言葉を優先的に選択することを特徴とする請求項11に記載の音声入力解釈方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、入力音声进行解釈する音声入力解釈装置及び音声入力解釈方法に関する。

【0002】

【従来の技術】近年、パーソナルコンピュータを含む計算機システムにおいて、従来のキーボードやマウスによる入力に加えて、音声情報を入力することが可能となってきた。

【0003】また、自然言語解析や自然言語生成、あるいは音声認識や音声合成技術あるいは対話処理技術の進歩などによって、利用者と音声入出力で対話する音声対話システムの要求が高まっており、自由発話による音声入力によって利用可能な対話システムである「TOSBURGER-II」(電子情報通信学会論文誌、Vol. 1J77-D-II、No. 8、pp. 1417-1428、1994)など、様々な音声対話システムの開発がなされている。

【0004】このような音声対話システムに利用される音声による入力方法は、特にキーボードのような習熟を

要するものではなく、誰にでも扱える入力方法であるので、誰もが利用する社会システム等への利用が期待され、より高度な音声処理技術への要求が高まっている。

【0005】従来、音声入力の解釈は、利用者から例えばマイクなどを通じて入力される音声入力を取り込み、例えば信号強度などによって音声分析単位の候補を推定し、分析単位項の例えばFFT(高速フーリエ変換)などを用いた分析によって特徴パターンなどを抽出し、あらかじめ用意した標準パターンと抽出パターンとを、例えば、複合類似度法、DP(ダイナミックプログラミング)法、あるいはHMM(隠れマルコフモデル)などを用いた照合を行い、入力された音声の認識を行い、音声認識結果に対して、構文解析、意味解析、などを行うことで利用者からの入力の意味内容や、発話意図を抽出することによって行われている。

【0006】従来、こういった音声対話システムなどにおける音声入力解釈方法において音声認識を行う際には、あらかじめ用意していた単語あるいは文章のパターンとの照合を行っていた。しかし、この方法では、利用者は発言できる単語あるいは文章(すなわちそのシステムが解釈可能な単語あるいは文章)を明確に記憶する必要があり、利用者負担を与えていた。

【0007】更に、利用者が、発言できる単語あるいは文章の一部のみを記憶している場合においても、利用者がその記憶されている一部分を入力しても、あらかじめ用意されていたパターンとは異なる音声入力とみなされ誤認識が生じ、結果として利用者の意図に反した動作を出力することが多く、利用者負担を与えていた。

【0008】例えば、社会システムの具体例として道案内のタスクを持つものを挙げると、利用者が知っている情報が「東京ステインホテル」の一部の「東京…ホテル」である場合に、そのホテルに関する情報を聞き出そうとして「東京なんとかホテル」と入力しても、あらかじめシステム中に準備された実在するホテルの名前のパターンとは異なるものであるため、誤認識が生じ、利用者の意図に反する情報が提示されるという結果となり、利用者にはなんの利益もなさないことになる。

【0009】また、利用者が、発言できる(あるいは当然にシステム中に登録されているものと期待される)単語あるいは文章のリズムのみを記憶しているような場合に、その単語あるいは文章のリズムのみを保有するような別の単語あるいは文章を入力しても、従来のシステムでは正式な入力として受け付けることができず、誤認識が生ずるため、利用者の意図した動作が行われることはなく、利用者負担を与えていた。

【0010】例えば、社会システムの具体例として上記と同様に道案内のタスクを持つものを挙げると、ある利用者が「丸の口ホテル」に関する情報を取得しようとする際に、この利用者が持っている情報が「丸の口ホテル」のリズムと一部の「…ホテル」である場合に、その

ホテルに関する情報を聞き出そうとして、「なんとかホテル」という意味で「ララララホテル」あるいは「ホニャララホテル」あるいは「タラララホテル」などと「丸の口ホテル」の持つリズムを意識して（あるいは真似て）適宜発声して入力しても、誤認識が生じ、利用者の意図に反する情報が提示されるという結果となり、利用者にはなんの利益もなさないことになる。

【0011】以上示したように、従来の音声入力解釈方法では、あらかじめ準備された単語あるいは文章のパターンでしか理解できないために、利用者に多大な負担を与えていた。

【0012】

【発明が解決しようとする課題】このように、音声入力を伴う装置において従来の音声入力解釈方法を適用すると、音声入力として受け付けられる単語あるいは文章のパターンがあらかじめ登録されているものに限定されているため、利用者が発声できる文章を明確に記憶する必要があり、利用者の負担が増加するという問題があった。

【0013】また、利用者が、発言できる単語あるいは文章の一部のみを記憶している場合においても、利用者がその記憶されている一部分を入力しても、あらかじめ用意されていたパターンとは異なる音声入力とみなされ誤認識が生じ、結果として利用者の意図に反した動作を出力することが多く、利用者の負担が増加するという問題があった。

【0014】また、利用者が、発言できる単語あるいは文章のリズムのみを記憶している場合においては、従来のシステムでは正式な入力として受け付けることができず、誤認識が生ずるため、利用者の意図した動作が行われることはなく、利用者の負担が増加するという問題があった。

【0015】本発明は、上記事象を考慮してなされたもので、利用者が正確に、発声できる単語あるいは文章を記憶しなくとも、アプリケーション部分が適切に動作するように解釈することのできる音声入力解釈装置を提供することを目的とする。

【0016】また、本発明は、利用者が発声可能な単語あるいは文章の一部分のみを記憶している場合でも音声の誤認識をおさえ、音声入力をもつシステムの出力を利用者の意図にそったものへと導くことのできる音声入力解釈装置を提供することを目的とする。

【0017】また、本発明は、利用者が発声可能な単語あるいは文章のリズムのみを記憶している場合でも音声の誤認識をおさえ、音声入力をもつシステムの出力を利用者の意図にそったものへと導くことのできる音声入力解釈装置及び音声入力解釈方法を提供することを目的とする。

【0018】

【課題を解決するための手段】本発明（請求項1）は、

入力音声进行解釈して該当する語彙の情報を出力する音声入力解釈装置において、正規の語彙に関する第1の情報、および該正規の語彙の一部が予め定められた代替表現に置き換えられて音声入力されることを考慮した該正規の語彙に関する第2の情報を記憶する手段と、入力音声を音声認識する手段と、前記第2の情報をもとに、前記音声認識結果から前記代替表現を検出する手段と、この手段により前記認識結果から前記代替表現が検出された場合、少なくとも前記入力音声の認識結果に含まれる該代替表現以外の語彙の部分をもとに、前記第1の情報を検索して、該当する語彙を求める手段とを備えたことを特徴とする。

【0019】好ましくは、前記該当する語彙が複数検索された場合、少なくとも前記代替表現に対応する音声の音韻的特徴に基づいて、該当する語彙の優先度を評価する手段をさらに備えるようにしてもよい。

【0020】本発明（請求項3）は、入力音声を解釈して該当する語彙の情報を出力する音声入力解釈装置において、任意の言葉の代替となる代替表現によって音声認識対象となる予め定められた正規の語彙の一部を代替した代替表現を語彙の一種として記憶する語彙記憶手段と、前記語彙記憶手段に記憶されている語彙のうち前記代替表現を含まない前記正規の語彙の表記および韻律情報を記憶する韻律情報記憶手段と、音声入力装置を介して入力された音声に対し、前記語彙記憶手段を参照して、音声認識および音声の韻律に関する分析を行う音声分析手段と、前記音声分析手段による前記入力された音声に対する前記音声認識の結果および前記韻律に関する解析の結果に基づき、前記韻律情報記憶手段を参照して、前記代替表現の部分を前記正規の語彙の部分で置換する置換表現照合手段とを備えたことを特徴とする。

【0021】本発明によれば、利用者が語彙記憶手段に記憶されている語彙を明確に覚えていなくとも、明確に覚えていない部分を代替表現を利用して音声入力を行うことができ、入力された代替表現に対応する適切な表現を検索し、代替表現を含まない適切な語彙に置換することが可能となる。

【0022】本発明（請求項4）は、音声入力装置から入力された音声を分析し、音声認識し、音声認識結果を含む音声分析結果を出力する手段と、該音声認識を行う際に認識対象となる語彙を記憶する語彙記憶手段とを備えた音声入力解釈装置において、任意の言葉の代替となる代替表現を記憶する代替表現記憶手段と、入力された音声情報から前記代替表現記憶手段に記憶されている語彙と同じ表現を検出する代替表現検出手段と、前記語彙記憶手段に記憶されている語彙をさらに分割して別単語としたものを記憶する置換表現記憶手段と、前記代替表現検出手段により前記代替表現の検出された入力音声情報における該代替表現でない部分の音声認識を、前記置換表現記憶手段に記憶されている語彙を音声認識対象と

して実行し、この音声認識結果を利用して前記置換表現記憶手段に記憶されている語彙から代替表現された言葉として妥当な語彙を検索する処理手段とを備えたことを特徴とする。

【0023】本発明によれば、利用者が語彙記憶手段に記憶されている語彙を明確に覚えていなくとも、明確に覚えていない部分を代替表現を利用して音声入力を行うことができ、また、任意の言葉の代替となる表現を音声入力から検出し、検出された代替表現に対応する適切な表現を検索することが可能となる。

【0024】好ましくは、前記処理手段は、前記音声認識を音節または音韻単位で行い、この音節または音韻単位の認識結果を参照することにより、前記代替表現の一部として前記正規の語彙の一部が付加されて発声された部分を検出し、前記置換表現記憶手段に記憶されている語彙から代替表現された表現を検索する際に、前記検出結果に適合した表現を優先的に選択するようにしてもよい。

【0025】これによって、利用者の代替表現の中に一部正しい発声をおりませた音声入力に対して、一部の正しい発声の情報に適合したより適切な表現を検索することができる。

【0026】好ましくは、前記代替表現検出手段は、入力音声の韻律について分析し、前記処理手段は、前記置換表現記憶手段に記憶されている語彙から代替表現された表現を検索する際に、前記分析の結果得られた韻律の条件に適合または近似した言葉を優先的に選択するようにしてもよい。

【0027】本発明（請求項7）は、入力音声进行解釈して該当する語彙の情報を出力する音声入力解釈方法において、入力音声を音声認識し、予め定められた正規の語彙の一部が予め定められた代替表現に置き換えられて音声入力されることを考慮した該正規の語彙に関する情報をもとに、前記音声認識結果から前記代替表現を検出し、前記認識結果から前記代替表現が検出された場合、少なくとも前記入力音声の認識結果に含まれる該代替表現以外の語彙の部分をもとに、予め定められた正規の語彙に関する情報を検索して、該当する語彙を求めることを特徴とする。

【0028】好ましくは、前記該当する語彙が複数検索された場合、少なくとも前記代替表現に対応する音声の音韻的特徴に基づいて、該当する語彙の優先度を評価するようにしてもよい。

【0029】本発明（請求項9）は、入力音声を解釈して該当する語彙の情報を出力する音声入力解釈方法において、音声入力装置を介して入力された音声に対し、任意の言葉の代替となる代替表現によって音声認識対象となる予め定められた正規の語彙の一部を代替した代替表現を語彙の一種として記憶する語彙記憶手段を参照して、音声認識および音声の韻律に関する分析を行い、前

記入力された音声に対する前記音声認識の結果および前記韻律に関する解析の結果に基づき、前記語彙記憶手段に記憶されている語彙のうち前記代替表現を含まない前記正規の語彙の表記および韻律情報を記憶する前記韻律情報記憶手段を参照して、前記代替表現の部分を前記正規の語彙の部分で置換することを特徴とする。

【0030】本発明（請求項10）は、入力音声を音声認識を通じて解釈し、該音声認識を行う際に認識対象となる語彙を記憶する語彙記憶手段のうちの該当する語彙の情報を出力する音声入力解釈方法において、入力された音声情報から、任意の言葉の代替となる代替表現を記憶する代替表現記憶手段に記憶されている語彙と同じ表現を検出し、前記代替表現の検出された入力音声情報における該代替表現でない部分の音声認識を、前記語彙記憶手段に記憶されている語彙をさらに分割して別単語としたものを記憶する置換表現記憶手段に記憶されている語彙を音声認識対象として実行し、この音声認識結果を利用して前記置換表現記憶手段に記憶されている語彙から代替表現された言葉として妥当な語彙を検索することを特徴とする。

【0031】好ましくは、前記語彙を検索するにあたっては、前記音声認識は音節または音韻単位で行い、この音節または音韻単位の認識結果を参照することにより、前記代替表現の一部として前記正規の語彙の一部が付加されて発声された部分を検出し、前記置換表現記憶手段に記憶されている語彙から代替表現された表現を検索する際に、前記検出結果に適合した表現を優先的に選択するようにしてもよい。

【0032】好ましくは、前記置換表現記憶手段に記憶されている語彙から代替表現された表現を検索する際に、入力音声の韻律について分析を行った結果得られた韻律の条件に適合または近似した言葉を優先的に選択するようにしてもよい。

【0033】本発明によれば、明確な表現の代替となるワイルドカード表現を検出する機能、またその代替された適切な表現を検索し、置換する機能を追加することによって、あるいは、ワイルドカード表現で実際に置換した語彙をもった語彙記憶手段を伴った音声分析機能と、またその代替された適切な表現を検索し、置換する機能を追加することによって、利用者が発声可能な語彙の一部しか記憶していない場合でも、ワイルドカード表現を用いた音声入力を受け入れることによって、その音声入力の解釈を行うことが可能となる。

【0034】また、本発明によれば、利用者が発声可能な語彙のリズムしか記憶していない場合でも、それに対応したワイルドカード表現を用いた音声入力を受け入れることによって、その音声入力の解釈を行うことが可能となる。

【0035】このように、本発明によれば、利用者が音声入力をもつ装置の許容する語彙を明確に覚えなくと

も、その音声入力を受け入れ、解釈することができる柔軟な音声入力解釈装置が構築できる等の実用上多大な効果が奏せられる。

#### 【0036】

【発明の実施の形態】以下、図面を参照しながら発明の実施の形態を説明する。

【0037】（第1の実施形態）まず、本発明の第1の実施形態について説明する。

【0038】図1に本実施形態に係る音声入力解釈装置の構成例を示す。図1に示されるように、本実施形態の音声入力解釈装置1は、音声分析部101、語彙記憶部102、置換表現照合部103、韻律情報記憶部104を備えている。なお、入力音声をアナログ信号からデジタル信号に変換するA/D変換器は、音声入力解釈装置1内に設けても、音声入力装置100側に設けてもよい。

【0039】音声分析部101は、置換表現照合部103と、語彙記憶部102と、マイクなどの音声入力装置100に接続し、例えば「パターンマッチング法による連続単語および連続音節の音声認識アルゴリズム」（電子情報通信学会論文誌、J-66-D、6、pp. 637-644）に開示されているような方式などで、語彙記憶部102に記録されている語彙を対象として、連続単語音声認識を行う。更に、例えば「ピッチパターン情報を利用したキーワードスポッティング」（日本音響学会講演論文集、平成8年9月、pp. 29-30）に開示されているような方式などにより、音声のピッチパターン情報などから解析を行い、韻律パラメータを生成する。そして、図4に示す情報を置換表現照合部103に渡す。尚、連続単語音声認識の方式や、韻律パラメータを生成する方式については、上記にあげた方式に限らず、その他の方式でも構わない。

【0040】語彙記憶部102は、音声分析部101に接続し、音声認識対象の語彙を記録する部分であり、正規の語彙のそれぞれについて図2に示すような情報を記憶するとともに（この場合、図2の情報においてワイルド表現に関するものは存在しない）、これに加えて、例えば「なんとか」あるいは「ホニャララ」などのような任意の数単語に置換される表現であるワイルドカード表現で正規の語彙の一部をワイルドカード表現で置換した語彙のそれぞれについて、図2に示すような情報を記憶する。

【0041】図2の情報の詳細については後述するが、「表象」情報の記述形式について先に触れておく。音声分析部101で行われる連続単語音声認識では、認識結果を複数の単語の連なりとして表現できるため、その単語同士の別れ目を記号“ / ”（スラッシュ）で表している。また、以下の説明でもこの単語同士の別れ目の表記には記号“ / ”を用いる。

【0042】また、使用されているワイルドカード表現

として、「なんとか」のようにいくつかの単語に置換されると考えられる表現である数単語置換語と、「ホニャララ」のようにその置換されるべき表現のリズムを表していると考えられるリズム語との一方または両方を定義しておく。使用する数単語置換語やリズム語の具体的内容やその種類数はシステムに応じて適宜定めてよい。

【0043】図3に「東京ステインホテル」とワイルドカード表現の数単語置換語「なんとか」とリズム語「ホニャララ」から生成される語彙の例を示す。これより、ワイルドカード表現が「東京」「ステイン」「ホテル」の中の数単語に置換されている語彙を生成し、また、特に「ホニャララ」のようなリズム語は置換される表現と等しい長さに拡張されて置換されている語彙を生成していることが分かる（この場合、「ラ」の数で長さを調整している）。

【0044】図2は語彙記憶部102で記録する情報の一覧である。併せて語彙「東京ホニャララホテル」の場合の例も示してある。「表象」情報は、その語彙の文字列を表す情報である。図2の例では「東京／ホニャララ／ホテル」と3単語連なった表象として記録されている。「ワイルドカード表現の有無」情報は、その語彙に先に述べたワイルドカード表現が含まれていたかどうかを表す情報である。この場合は「ホニャララ」がワイルドカード表現にあたるので「有り」が記録されている。「表現の種類」情報は、その語彙に含まれる単語のそれぞれがワイルドカード表現か、ワイルドカード表現ではない非ワイルドカード表現かを表す情報である。ワイルドカード表現の単語には「代替」を、非ワイルドカード表現には「確定」を与える。この例では、単語「東京」「ホテル」が非ワイルドカード表現で、「ホニャララ」がワイルドカード表現であるので、（確定／代替／確定）と情報が与えられている。「ワイルドカード表現の種類」情報は、その語彙に含まれているワイルドカード表現が、数単語置換語か、リズム語かを表す情報である。この場合は「ホニャララ」がリズム語と定義されているので「リズム語」と記録している。「音声認識パラメータ」情報は音声分析部101で行われる音声認識のために必要に応じてパラメータを記述するものである（なお、ここで使用する音声認識方式は本発明の本質ではないのでこのパラメータについての詳細な説明は省略する）。

【0045】図4は音声分析部101から置換表現照合部103へ渡される情報の一覧である。併せて、「東京ホニャララホテル」と入力された場合の例も示してある。「認識結果」情報は、音声分析部101で連続単語認識された結果の表象を表す情報である。図4の例では入力された音声信号の認識結果として、「東京／ホニャララ／ホテル」と示されている。「単語発声時間」情報は、音声分析部101で連続単語認識された際に得られる、各単語の発声時間を表す情報である。この例では



(650msec/820msec/510msec)と示されているが、これらの数字は順に「東京」「ホニャララ」「ホテル」に対応している発声時間を表している。「韻律パラメータ」情報は、音声分析部101で解析された韻律パラメータを表す情報である。この情報は、韻律パラメータの解析手段によって形態が異なるものとなるが、ここでは、イントネーションあるいは基本周波数の時間的推移を用いた場合を示す。そして、ここでは、得られるであろう韻律パラメータを模式的に表している。図4で使用されている矢印記号「→」はその言葉の抑揚を模式的に表現しており、上方にある矢印が抑揚の高い部分を、下方にある矢印が抑揚の低い部分を表している。「ワイルドカード表現の有無」情報は、入力された音声に先に述べたワイルドカード表現が含まれていたかどうかを表す情報である。この場合は「ホニャララ」がワイルドカード表現にあたるので「有り」が出力されている。「表現の種類」情報は、その語彙に含まれる単語のそれぞれがワイルドカード表現か、ワイルドカード表現ではない非ワイルドカード表現かを表す情報である。この情報は対応する語彙に関する図2の「表現の種類」情報を参照すれば得られ、また、その表記方法は図2における「表現の種類」情報と同じである。図4の例では、単語「東京」「ホテル」が非ワイルドカード表現で、「ホニャララ」がワイルドカード表現であるので、(確定/代替/確定)と情報が与えられている。「ワイルドカード表現の種類」情報は、入力音声に含まれていたワイルドカード表現が数単語置換語かリズム語かを識別するための情報である。この場合は「ホニャララ」がリズム語であるので「リズム語」を出力している。

【0046】置換表現照合部103は、音声分析部101と、韻律情報記憶部104に接続し、ワイルドカード表現が検出された場合に、そのワイルドカード表現部分に対応する適切な表現を照合する。この部分の詳細については後述する。

【0047】韻律情報記憶部104は、置換表現照合部103に接続し、語彙記憶部102に登録されている語彙のうち、ワイルドカード表現を含まない正規の語彙について、図5に示すような情報を記録する。

【0048】図5は韻律情報記憶部104で記録されている情報の一覧である。また、あわせて「東京ステインホテル」の例も示している。「表象」情報はその語彙の表象情報である。「標準時間」情報は記録されている言葉のサンプルの発声時間を表している。その語彙が連続単語として分離できる場合には、そのそれぞれの単語の発声時間を記録しておく。この情報の表記方法は図4の「単語発声時間」情報のそれと同じである。「韻律」情報は記録されている言葉のサンプルから解析される韻律情報を表している。但し、韻律情報を解析する方法は音声分析部101で行っている方法と同じ方法でなければ

ならない。また、韻律情報記憶部104から出力される韻律情報も、音声分析部101から置換表現照合部103へ渡す韻律パラメータ情報と同形式のものでなければならない。図5の例は、図4と同様に解析後得られるであろう韻律情報を模式的に表している。

【0049】図6は本実施形態で重要な働きをする置換表現照合部103の動作のフローチャートである。以下、図6を参照して、処理の流れを説明する。

【0050】(ステップS101)ここでは、音声分析部101の音声認識結果にワイルドカード表現があるかどうかを確認する。これは、音声分析部101から渡される図4の「ワイルドカード表現の有無」情報で確認が可能である。そして、ワイルドカード表現が存在する場合はステップS102へ、ワイルドカード表現が存在しない場合は認識結果を出力し、処理を終了する。

【0051】(ステップS102)このステップでは、渡された音声認識結果に適合しかつ出力対象となる語彙を韻律情報記憶部104から選択する。例えば、韻律情報記憶部104に記録されている情報(図5)の表象情報を利用し、音声認識結果に含まれている非ワイルドカード表現部分を音声分析部101から渡される情報(図4)の表現の種類情報を参照することにより求め、その非ワイルドカード表現の存在位置条件に適合する語彙を、ワイルドカード表現は1単語以上の長さを持つものと考え、非ワイルドカード表現部分を条件とすることにより、選択する。

【0052】例えば、得られた音声認識結果が「東京/なんとか/ホテル」、表現の種類情報が(確定/代替/確定)であったとすると、「東京/ステイン/ホテル」、「東京/エンター/コンチネンタル/ホテル」などのように、単語「東京」が最初に存在し、かつ、単語「ホテル」が最後に存在し、かつ、その間に少なくとも1単語以上存在するものを適合する語彙として選択する。

【0053】(ステップS103)このステップでは、音声認識結果に含まれるワイルドカード表現が数単語置換語かリズム語かを判別する。これは、音声分析部101から渡される図4の「ワイルドカード表現の種類」情報で確認が可能である。そして、リズム語の場合はステップS104へ、数単語置換語の場合はステップS102で抽出された語彙を出力し、処理を終了する。尚、出力する語彙が複数存在する場合は、その中のいくつかを出力しても、全てを出力してもよく、複数個の解を利用者に提示して選択させるなどの処理は、出力先のアプリケーション特有の処理(図中200)で決定される。

【0054】(ステップS104)このステップでは、ステップS102で抽出された語彙について、音声分析部101から渡された「単語発声時間」情報と韻律情報記憶部104に記録されている「標準時間」情報とを比較することによって、更に出力語彙を限定する。例え

ば、「東京／ホニャララ／ホテル」の場合は非ワイルドカード表現部分である「東京」「ホテル」の発声時間と、対象語彙に関する韻律情報記憶部104の標準時間情報に記録されている「東京」「ホテル」の標準時間との比率をそれぞれ計算し、その比率の平均値で入力信号のワイルドカード部分の発声時間を伸長し、伸長された発声時間と標準時間とを比較し、あるしきい値以内のもののみを抽出する。尚、この処理で語彙を限定しない場合は、時間を比較することによって、出力する語彙の優先順位を決定することも可能である。

【0055】(ステップS105)このステップでは、ステップS104で抽出された語彙について、音声分析部101から渡された「韻律パラメータ」情報と韻律情報記憶部104に記録されている「韻律パラメータ」情報とを比較することによって、出力する語彙を決定する。例えば、「ピッチパターン情報を利用したキーワードスポッティング」(日本音響学会講演論文集、平成8年9月、pp. 29-30)に開示された方法により、DP法を利用したマッチングを行うことによって比較を行う。尚、この比較方法は構成される韻律パラメータによっても異なるが、本実施形態では、構成されるパラメータを利用できるものであれば、任意の韻律比較方法を利用しても構わない。そして、発声した音声に最も韻律情報が類似している語彙を出力し、処理を終了する。あるいは、複数候補存在する場合には、韻律情報が類似している順に優先順位をつけて出力しても良い。

【0056】以上が、本発明に係る置換表現照合部103の構成とその機能、および処理方法である。

【0057】続いて、上述した音声入力解釈方法について、更に詳しく説明する。ここでは、アプリケーションとして地図情報システムとして利用者が音声入力を行った場合の働きを具体例として説明を行う。

【0058】この地図情報システムには4つのホテルの情報(パルスホテル、東京ステインホテル、東京丸の口ホテル、東京エンターコンチネンタルホテル)が登録されており、その4つのホテルの名称が語彙記憶部102に記録されているとする。また、語彙記憶部102にはワイルドカード表現として前述したリズム語「ホニャララ」が登録されており、上記の4つのホテルと「ホニャララ」から生成される語彙を合わせて、語彙記憶部102には図7に示した語彙が登録されているとする。

【0059】また、韻律情報記憶部104には登録された4つのホテルの名称から表象情報、韻律情報、標準時間情報を求めることによって、図8に示すような情報が記録されているとする。

【0060】そして、利用者が「東京ステインホテル」について聞きたいが、「ステイン」の部分を確認に記憶していなかったとし、この地図情報システムに「トウキョウホニャララホテル」という音声入力が行われたものとする。ただし、この発言に含まれるワイル

ドカード表現「ホニャララ」は「ステイン」のリズムを意識した発言とする。

【0061】以下、本具体例の場合における各部の動きについて述べる。

【0062】まず、音声分析部101では、入力された音声に対して図7にある語彙で連続単語認識を実行する。そして、認識結果として「東京／ホニャララ／ホテル」が選択されたとし、認識処理時に得られる発声時間情報と、入力音声から抽出される韻律情報と合わせて、図9に示す情報を置換表現照合部103に出力する。

【0063】この情報を受けた置換表現照合部103は以下のような処理を行う。

【0064】(ステップS101)渡されたワイルドカード表現の有無情報から、認識結果にワイルドカード表現があると判断して、ステップS102に進む。

【0065】(ステップS102)認識結果情報「東京／ホニャララ／ホテル」と、表現の種類情報(確定／代替／確定)とから、非ワイルドカード表現を「東京」「ホテル」とし、これら2単語の存在位置条件に適合するものを韻律情報記憶部104に登録された語彙(図8)から検索する。この場合は、最初に「東京」、最後に「ホテル」があり、その間に少なくとも一単語存在する語彙が検索条件に当てはまるとする。そして、「東京ステインホテル」、「東京丸の口ホテル」、「東京エンターコンチネンタルホテル」が検索され、「パルスホテル」は出力候補から外されるか、下位の候補とされる。

【0066】(ステップS103)渡されたワイルドカード表現の種類情報から、ワイルドカード表現「ホニャララ」はリズム語であるとして、ステップS104に進む。

【0067】(ステップS104)ステップS102で選択された語彙から、まず「東京ステインホテル」から標準時間情報(図8)と、音声分析部101から渡された単語発声時間情報とを比較する。例えば、まず、非ワイルドカード表現である「東京」、「ホテル」に関する両者の比率(標準時間情報／単語発声時間情報)を計算すると、「東京」: $700/650=1.0769$ 、「ホテル」: $550/510=1.0784$ となる。次に、これらの比率の平均を計算し、その結果得られる数値(1.0777)を入力時間を韻律情報記憶部104にある標準時間と同スケールとする伸長係数とする。そして、ワイルドカード表現にあたる「ホニャララ」部分を伸長した後のワイルドカード表現部の入力時間は $820\text{ msec} \times 1.0777 = 884\text{ msec}$ となる。次に、「東京」と「ホテル」の間にあり、ワイルドカード表現で代替されたと考えられる「ステイン」部分の標準時間は $900\text{ msec}$ となる。そして、これら2つの入力時間を比較する(例えばしきい値処理)ことによ

って、ワイルドカード表現部分の時間的整合がとれているかを調べる。図10にステップS102で選択された語彙に関して、上記の計算を行った結果を示す。

【0068】ここで、「東京エンターコンチネンタルホテル」に関しては、「ホニャララ」で「エンター／コンチネンタル」を代替表現したものと考えられるので、標準時間(700msec/650msec/1050msec/550msec)の内、「エンター」「コンチネンタル」に相当する650+1050=1700msecがワイルドカード表現「ホニャララ」に対応する標準時間である。そして、例えば、伸長後の時間と標準時間との差を計算し、その絶対値があるしきい値より大きいものは出力候補から外す処理を行うとし、そのしきい値を100msecとすると、上記の表より「東京エンターコンチネンタルホテル」が出力候補から外されるか、下位の候補とされる。

【0069】(ステップS105)これまでの処理によって外されなかった「東京ステインホテル」「東京丸の口ホテル」についてその韻律情報のマッチングを行う。そして、その結果、音声分析部101から渡された韻律情報と近い韻律情報をもつ語彙が出力されるか、あるいは優先順位の高い語彙となる。ここで、「東京ステインホテル」の韻律情報の方が入力音声の韻律と近いものと判断され、優先順位の高い語彙として「東京ステインホテル」を出力し、出力先のアプリケーション特有の処理(図中200)で適切な処理を行う。また、アプリケーション特有の処理が複数候補に対して処理を行うことが可能であれば、下位の候補として「東京丸の口ホテル」を、必要ならば、更に下位の候補として順に「東京エンターコンチネンタルホテル」、「パルスホテル」も併せて出力する。

【0070】以上で「東京ホニャララホテル」と音声入力された場合の処理を終了する。

【0071】以上の説明によって、本実施形態に係る音声入力解釈装置は、利用者が「東京ステインホテル」という名称を明確に記憶していない状態でも、分からない部分をワイルドカード表現を用いて、「東京ホニャララホテル」と音声入力することによって、適切な名称に解釈してアプリケーション部分に情報を出力することが可能であり、また、利用者が知っていても文字列には表せないリズムでの表現をワイルドカード表現を利用して「東京ホニャララホテル」と入力し、本実施形態に係るシステムがその発声時間情報、韻律情報を解釈することにより、同じく「東京…ホテル」の形式の名称を持つ「東京エンターコンチネンタルホテル」、「東京丸の口ホテル」よりも、「東京ステインホテル」のほうが優先され、利用者の入力した音声情報が有効に利用されていることがわかる。

【0072】(第2の実施形態)次に、本発明の第2の実施形態について説明する。

【0073】第1の実施形態では音声認識方式として連続単語認識を用いるものであったが、本実施形態は音声認識方式が連続単語認識でなくとも適用可能としたものである。

【0074】図11に本実施形態に係る音声入力解釈装置の構成例を示す。図11に示されるように、本実施形態の音声入力解釈装置2は、音声分析部201、語彙記憶部202、ワイルドカード表現検出部203、ワイルドカード表現記憶部204、置換表現照合部205、置換表現記憶部206を備えている。なお、入力音声をアナログ信号からデジタル信号に変換するA/D変換器は、音声入力解釈装置2内に設けても、音声入力装置100側に設けてもよい。

【0075】音声分析部201は、置換表現照合部205と、語彙記憶部202と、置換表現記憶部206に接続し、置換表現照合部205から音声認識要求が来ると、語彙記憶部202か置換表現記憶部206のどちらか一方の指定された語彙を用いて音声単語認識を行い、その結果を置換表現照合部205に出力する。また、認識方法の要求に応じて単音節認識を行い、認識結果をモーラ記号列として置換表現照合部205に出力する。なお、これらの音声認識方法については本発明の本質ではないので、これらについての詳細な説明は省略する。

【0076】語彙記憶部202は、音声分析部201と、置換表現照合部205とに接続し、音声認識対象の(正規の)語彙を記録する部分であり、音声認識対象の各語彙について図12に示す情報を音声分析部201、置換表現照合部205が参照・利用可能な形式で記録する。

【0077】図12は語彙記憶部202が記録する情報の一覧である。併せて、語彙「東京ステインホテル」に対応して語彙記憶部202が記録する情報を例として示す。「表象文字列」情報は、登録する語彙を表す文字列である。「モーラ記号列」情報は、表象文字列の読みをモーラ記号列で記述したものである。「モーラ記号列長」情報は、モーラ記号列情報で記録されたモーラ記号列のモーラ記号の数を表している。「韻律パラメータ」情報は、例えば「ピッチパタン情報を利用したキーワードスポッティング」(日本音響学会講演論文集、平成8年9月、pp. 29-30)に開示された方式などにより、音声のピッチパタン情報などから解析を行い、構成される韻律パラメータを記録する。尚、韻律パラメータを生成する方式については上記の方式に限らず、その他の方式であっても構わない。また、図12の例では、得られるであろう韻律情報を模式的に表している。この表記方法は第1の実施形態のものと同様である。「音声認識に必要なパラメータ」情報は、本発明を実施する際に音声分析部201で使用する音声認識のために必要に応じてパラメータを記述するものである(なお、ここで使用する音声認識方式は本発明の本質ではないのでこのバ

ラメータについての詳細な説明は省略する)。

【0078】ワイルドカード表現記憶部204は、ワイルドカード表現検出部203に接続し、例えば「なんとか」あるいは「ホニャララ」などのような任意の数単語に置換される表現であるワイルドカード表現を、ワイルドカード表現検出部203が参照・利用可能な形式で記憶する。また、記憶するワイルドカード表現を「なんとか」「なになに」等の数単語に置換される表現の数単語置換語と、「ホニャララ」「タラララ」等の置換されるべき表現のリズムを表すリズム語とに分けて記憶する。

【0079】ワイルドカード表現検出部203は、マイクなどの音声入力装置100と、ワイルドカード表現記憶部204と、置換表現照合部205に接続し、ワイルドカード表現記憶部204に記憶されているワイルドカード表現の語彙を例えば「ワードスポッティングによる音声認識における雑音免疫学習」(電子情報通信学会論文誌Vol. J-74-D-II 1991年2月pp. 121-129)に開示されている方法などを用いて検出する。尚、特定の語彙を検出できる手法であれば、上記の方式に限らず、他の検出方式でも構わない。そして、ワイルドカード表現検出部203は、図13に示したような情報を置換表現照合部205に与え、処理を渡す。

【0080】図13はワイルドカード表現検出部203から置換表現照合部205に渡す情報の一覧である。また、併せて「トウキョウホニャララホテル」と入力された場合の例も示す。「ワイルドカード表現の有無」はワイルドカード表現がワイルドカード表現検出部203で検出されたかどうかを表す情報である。この例では「ホニャララ」がワイルドカード表現にあたり、「有り」を出力している。「原信号」は音声入力された元の信号であるが、ワイルドカード表現が検出された場合はそのワイルドカード表現の部分で切り離して置換表現照合部205に渡す。例では入力「トウキョウホニャララホテル」がワイルドカード表現「ホニャララ」で分離され「トウキョウ」「ホニャララ」「ホテル」と3つに分離されて順に置換表現照合部205に渡される。「ワイルドカード表現の位置」はワイルドカード表現が存在する場合に、切り離された原信号の何番目の信号がワイルドカード表現であるかを数値で表したものである。この例では3つに分離された原信号の2番めに「ホニャララ」があるので2が出力されている。「ワイルドカード表現の種類」は検出されたワイルドカード表現が、数単語置換語か、リズム語かを表す情報である。この例では「ホニャララ」をリズム語としている。これはワイルドカード表現記憶部204に登録されている情報によって異なる。「ワイルドカード表現のモーラ記号列長」は検出されたワイルドカード表現がリズム語であった場合にそのモーラ記号数を表す情報である。この例ではワイルドカード表現「ホニャララ」はモーラ記号数4である。「ワ

イルドカード表現の韻律情報」は検出されたワイルドカード表現がリズム語であった場合にその韻律を表す情報である。これは、入力された音声のビッチパターン情報などから解析を行い、置換表現照合部205に渡される。尚、韻律パラメータを生成する方式については、生成される韻律パラメータが、語彙記憶部202、置換表現記憶部206に登録される形式と同じものになるものでなければならない。

【0081】置換表現照合部205は、ワイルドカード表現検出部203と、音声分析部201と、語彙記憶部202と、置換表現記憶部206に接続し、ワイルドカード表現が検出された場合に、そのワイルドカード表現部分に対応する適切な表現を照合する。この部分の詳細について後述する。

【0082】置換表現記憶部206は、音声分析部201と、置換表現照合部205に接続し、語彙記憶部202に登録されている語彙から、例えば「東京ステインホテル」から「東京」「ステイン」「ホテル」のように、更に単語として意味のあるものに分離することによって生成される単語、あるいは「東京ステイン」「ステインホテル」のように連続している単語の組合せとなる言葉を語彙記憶部202と同じ形式(図12参照)で記憶する。

【0083】図14は本実施形態で重要な働きをする置換表現照合部205の動作の概略構成である。以下、図14を参照して処理の流れを説明する。

【0084】(ステップS201)入力された音声入力にワイルドカード表現があるかどうか確認する。これはワイルドカード表現検出部203から与えられる「ワイルドカード表現の有無」情報(図13)から確認できる。そして、ワイルドカード表現が存在すればステップS204へ、ワイルドカード表現が存在しなければステップS202へ進む。

【0085】(ステップS202)ワイルドカード表現がないと判断された場合、そのまま音声認識処理を行う。入力された原信号に対して音声分析部201に語彙記憶部202に記憶されている語彙での単語認識を依頼する。

【0086】(ステップS203)音声分析部201から出力された音声認識結果をアプリケーション特有の処理(図中200)、あるいはより高等な音声分析処理に引渡し、処理を終了する。

【0087】(ステップS204)ワイルドカード表現があると判断された場合、ワイルドカード表現ではない部分がどのように発声、入力されたかを調べる。例えば、ワイルドカード表現検出部203から切り離されて渡される原信号と「ワイルドカード表現の位置」情報から、ワイルドカード表現の信号を求め、ワイルドカード表現ではない部分(非ワイルドカード表現部)の信号に対して、音声分析部201に置換表現記憶部206に記

憶されている語彙での単語認識を依頼する。以下の説明では、このステップで得られた音声認識結果を「部分認識結果」と呼ぶ。また、置換表現記憶部206に適切な語彙が存在しない場合は、その非ワイルドカード表現部に対応する部分認識結果は存在しないこととする。

【0088】(ステップS205)ここでは、非ワイルドカード表現部とワイルドカード表現との間に何らかの情報がどうかを調べる。これは、利用者が明確な単語の発音を知らない場合においても、始点終点の一部のみを知っている場合に「『ス』なんとか」のようにワイルドカード表現の前後に付与する形式で発声される場合にも対応するために行う。「『ス』なんとか」のように発声されると、ワイルドカード表現記憶部204に登録されているワイルドカード表現に「すなんとか」が登録されていなければ、ワイルドカード表現検出部203によって検出されるワイルドカード表現は「なんとか」であるので、利用者がワイルドカード表現を意図として発声した『ス』は非ワイルドカード表現の一部として処理されてしまう。このような場合においても、非ワイルドカード表現部の中に、ワイルドカード表現の一部とされた部分が存在するかどうかを判定し、存在する場合はワイルドカード表現の一部として処理できるようにするものである。

【0089】まず、検出されたワイルドカード表現部のそれぞれにモーラ記号列を記憶するバッファを準備する。このバッファは非ワイルドカード表現部とワイルドカード表現部の間に情報が検出できた場合に、その情報をモーラ記号で記憶するものである。また、検出された情報がワイルドカード表現部の前、後に現れる場合があるので、それに対応してバッファは各ワイルドカード表現につき、2つずつ準備される。ステップS205では検出された非ワイルドカード表現部のうち、ワイルドカード表現部に隣接している部分のそれぞれに対して、バッファに入力するモーラ記号を抽出する処理を行う。図15は検出された非ワイルドカード表現部の一つに対する処理(ステップS205の処理)の概略構成を示している。以下では、図15を参照しながら説明を行う。

【0090】(ステップS205-1)このステップでは、非ワイルドカード表現がどのように発声されているのかということを調べる。例えば、対象となった非ワイルドカード表現部に対して、音声分析部201に音節単位の音声認識を依頼する。以下の説明では、このステップで出力されてきたモーラ記号列を「部分音節認識結果」と呼ぶ。

【0091】(ステップS205-2)このステップでは、現在対象となっている非ワイルドカード表現に対応する部分認識結果が存在するかどうか確認する。その結果、部分認識結果が存在しない場合はステップS205-3へ、部分認識結果が存在している場合はステップS205-4へ進む。

【0092】(ステップS205-3)現在対象となっている非ワイルドカード表現部に対応する部分認識結果が存在しない場合、この非ワイルドカード表現部は置換表現記憶部206の語彙よりも短い表現をしていると判断できる。そこで、この非ワイルドカード表現部全てが隣接しているワイルドカード表現部の一部であるとし、この非ワイルドカード表現が隣接しているワイルドカード表現部の前部にあるか、後部にあるかを判定し、そのモーラ記号(列)を対応するバッファに記憶し、ワイルドカード表現部がリズム語である場合は、ワイルドカード表現検出部203から受けとった「ワイルドカード表現の文字列長」情報にバッファに記憶したモーラ記号数分だけ加え、この非ワイルドカード表現部がワイルドカード表現部の前部に存在する場合は、「ワイルドカード表現の位置」情報を1減少させて、終了する。

【0093】(ステップS205-4)ここでは、対象となっている非ワイルドカード表現部の中に、対応する部分認識結果の他に発音された言葉が含まれているかを確認する。例えば、現在対象となっている非ワイルドカード表現に対応する部分音節認識結果のモーラ記号列長と、部分認識結果のモーラ記号列長とを比較する。その結果、部分認識結果のモーラ記号列長の方が長い場合や、両者に等しい場合であればステップS205-5に進み、部分音節認識結果のモーラ記号列長の方が長い場合はステップS205-6に進む。

【0094】(ステップS205-5)現在対象となっている非ワイルドカード表現には対応する部分認識結果以上の情報はないと判断し、バッファには何も入力せず終了する。

【0095】(ステップS205-6)現在対象としている非ワイルドカード表現部に対応する部分認識結果のモーラ記号列長が同じ部分の部分音節認識結果のモーラ記号列長より短いので、現在対象となっている非ワイルドカード表現には対応する部分認識結果の他にワイルドカード表現の一部が発声されている可能性があるとして判断し、部分認識結果が非ワイルドカード表現部の原信号のどの部分に当たるのかを調べる。

【0096】例えば、図16のように部分認識結果のモーラ記号列を部分音節認識結果のモーラ記号列に逐次当てはめ、両者のモーラ記号列を比較することにより求める。図16では部分認識結果「東京」(モーラ記号列「トオキョオ」と)と部分音節認識結果「トオキョオス」とを比較しているが、部分認識結果のモーラ記号列長は4で、部分音節認識結果のモーラ記号列長は5となっており、部分認識結果は当てはめを開始する記号を「トオキョオス」の「ト」と「オ」(最初のオ)とする2つのパターンが考えられる。部分認識結果が更に短い場合は「キョ」以降を開始とするパターンが現れる。

【0097】そして、どの当てはめのパターンが最適かを判断し、「余り」の部分が何処かを決定する。この

「余り」の部分は、非ワイルドカード表現の中に含まれているワイルドカード表現部分の一部とすべき箇所であると考えられる。余り部分の決定方法としては、例えば、当てはめるときに一致したモーラ記号数を基準とする場合は、最も一致するモーラ記号数の多い場所を部分認識結果が存在する場所として選択し、部分認識結果のモーラ記号列が当てはまらない部分を「余り」として抽出する。

【0098】図16の例では最後の文字「ス」が余りの部分として抽出される。

【0099】また、一致するモーラ記号数が最大になるパターンが2種類以上存在するなど、部分認識結果の位置が一意に決定できない場合には、余りの部分は存在しないと判断する。あるいは、一致するモーラ記号数があるしきい値以下であった場合も余りの部分は存在しないと判断しても良い。

【0100】(ステップS205-7)ここでは、ステップS205-6の結果、余りの部分が抽出できたかどうかを確認する。余りの部分が存在していればステップS205-8へ、余りの部分が存在しなければステップS205-5へ進む。

【0101】(ステップS205-8)ここでは、ステップS205-6の結果、抽出された余りの部分がワイルドカード表現部分に隣接したところに存在するかどうかを確認する。図16の例ではワイルドカード表現「ナントカ」の直前に余りの部分「ス」が存在するので、「ナントカ」に隣接した場所に余り「ス」が存在すると判断される。逆に、「ナントカ」が「トウキョウス」の前に存在している場合は、「トウキョウス」の最後部分にある余り「ス」は「ナントカ」とは隣接していないと判断される。この場合は、余りが「ト」「トウ」などであればワイルドカード表現「ナントカ」の直後に余りが存在すると判断できる。隣接部分に余りが存在すれば、ステップS205-9へ進む。隣接部分に余りが存在しなければ、ステップS205-5へ進む。

【0102】(ステップS205-9)ここでは、ステップS205-6で抽出された余りの部分がワイルドカード表現に隣接しているので、この抽出された余りの部分が隣接しているワイルドカード表現部の一部であるとし、この余りの部分が隣接しているワイルドカード表現部の前部にあるか、後部にあるかを判定し、そのモーラ記号(列)を対応するバッファに記憶し、ワイルドカード表現部がリズム語である場合は、ワイルドカード表現検出部203から受けとった「ワイルドカード表現の文字列長」情報にバッファに記憶したモーラ記号数だけ加え、終了する。

【0103】上記の方法の他にも、音声分析部201にワイルドカード表現検出部203にあるような単語検出能力を付与すれば、切り離された原信号の中からステップS204で得られた音声認識結果の単語の検出を行

い、その後ワイルドカード表現との境の部分に余った信号を切りとり、音声分析部201に音節単位の音声認識を依頼することによって、上記と同じくワイルドカード表現の一部となるモーラ記号(列)を推定することも可能である。

【0104】(ステップS206)ステップS204～S205での処理によって得られた情報と、ワイルドカード表現検出部203から得られる図13に示した情報を検索条件として、語彙記憶部202に記憶されている語彙に一致するように、置換表現記憶部206に記憶されている語彙からワイルドカード表現部分にあてはまる言葉を検索する。図17はステップS206で行う動作のフローチャートである。以下、図17を参照して、処理の流れを説明する。

【0105】(ステップS206-1)このステップでは、渡された音声認識結果に適合しかつ出力対象となる語彙を語彙記憶部202から選択する。例えば、語彙記憶部202に記録されている情報(図12)の表表情報を利用し、非ワイルドカード表現部分の部分認識結果の存在位置条件に適合する語彙を、ワイルドカード表現は1単語以上の長さを持つものと考え、非ワイルドカード表現部分を条件とすることにより、選択する。そして、置換表現記憶部206に記録されている表現から、選択された語彙のワイルドカード表現で代替表現された部分を検索する。例えば、得られた部分認識結果が「東京」と「ホテル」で、更に、切り分けられた原信号の並びと、ワイルドカード表現の位置情報から「東京(ワイルドカード表現)ホテル」の順であると分かるとすると、「東京ステインホテル」などのように、単語「東京」が最初に存在し、かつ、単語「ホテル」が最後に存在し、かつ、その間に少なくとも1単語以上存在するものを適合する語彙として選択する。そして、置換表現記憶部206からワイルドカード表現された部分として、表現「ステイン」などが検索される。

【0106】(ステップS206-2)このステップではステップS205で処理されたバッファに記録されたモーラ記号(列)を検索条件として、ステップS206-1で抽出された表現から更に限定を行う。

【0107】(ステップS206-3)このステップでは、音声認識結果に含まれるワイルドカード表現が数単語置換語カリズム語かを判別する。これは、ワイルドカード表現検出部203から渡される図13の「ワイルドカード表現の種類」情報で確認が可能である。そして、リズム語の場合はステップS206-4へ進み、数単語置換語の場合はステップS206-2で抽出された表現と、部分認識結果からなる正規の語彙を出力し、処理を終了する。尚、出力する語彙が複数存在する場合は、その中のいくつかを出力しても、全てを出力してもよく、複数の解の利用者に提示して選択させるなどの処理は、出力先のアプリケーション特有の処理(図中20



0)で決定される。

【0108】(ステップS206-4)このステップでは、ステップS206-1で抽出された語彙について、ワイルドカード表現検出部203から渡されたワイルドカード表現のモーラ記号列長情報と置換表現記憶部206に記録されているモーラ記号列長情報とを比較することによって、更に出力語彙を限定する。例えば、両者のモーラ記号列長の差があるしきい値以内のもののみを抽出する。尚、この処理で語彙を限定しない場合は、出力する語彙の優先順位を決定することも可能である。

【0109】(ステップS206-5)このステップでは、ステップS206-4で抽出された語彙について、音声分析部101から渡された「韻律パラメータ」情報と韻律情報記憶部104に記録されている「韻律パラメータ」情報とを比較することによって、出力する語彙を決定する。例えば、「ピッチパターン情報を利用したキーワードスポッティング」(日本音響学会講演論文集、平成8年9月、pp.29-30)に開示された方法などにより、DP法を利用したマッチングを行うことによって比較を行う。尚、この比較方法は構成される韻律パラメータによっても異なるが、本実施形態では、構成されるパラメータを利用できるものであれば、任意の韻律比較方法を利用しても構わない。そして、発声した音声に最も韻律情報が類似している語彙を出力し、処理を終了する。あるいは、複数候補存在する場合には、韻律情報が類似している順に優先順位をつけて出力しても良い。

【0110】以上が本実施形態に係る置換表現照合部205の構成とその機能、および処理方法である。

【0111】続いて、上述した音声入力解釈方法について、更に詳しく説明する。ここでは、第1の実施例の説明の際に利用した地図情報システムの例を挙げ、利用者が音声入力を行った場合の働きを具体例として説明を行う。

【0112】この地図情報システムには東京駅周辺の4つのホテル(東京ステインホテル、東京丸の口ホテル、パルスホテル、東京エンターコンチネンタルホテル)が登録されており、その4つのホテルに関して、図18に示す情報と、それぞれの音声認識に必要なパラメータとが語彙記憶部202に記録されている。そして、置換表現記憶部206に登録される、これらの語彙から分離した単語および連続している単語の組合せとなる表現は図19に示すようになる。

【0113】また、ワイルドカード表現として数単語置換語「ナントカ」がワイルドカード表現記憶部204に登録されているとする。

【0114】次に、利用者が「東京ステインホテル」について聞きたいが、「ステイン」の部分明確に記憶していなかったとし、この地図情報システムに「トウキョウスナントカホテル」という音声入力が行われたものとする。

【0115】以下、表記を明確にするため、音声認識結果を得る前の波形信号を「シンゴウ」のように「…」で、音声認識結果を得た後に得られる文字列を「文字列」のように「…」で表す。

【0116】その入力を受け、まずワイルドカード表現検出部203においてワイルドカード表現の検出が行われる。信号「トウキョウスナントカホテル」にはワイルドカード表現「ナントカ」が含まれており、これがワイルドカード表現として検出される。そして、置換表現照合部205に図20のような情報が渡される。

【0117】以下は置換表現照合部205での処理である。

【0118】(ステップS201)ワイルドカード表現の有無情報からワイルドカード表現が存在することが確認される。

【0119】(ステップS204)分離されて渡される原信号と、ワイルドカード表現の位置情報から音声認識が必要な部分が信号「トウキョウス」「ホテル」であるとわかる。そして、音声分析部201にこの2つの信号の置換表現記憶部206に記録されている語彙セットでの単語認識を依頼する。その結果、部分認識結果として信号「トウキョウス」の認識結果が「東京」、「ホテル」の認識結果が「ホテル」と得られたとする。

【0120】(ステップS205)まず、信号「トウキョウス」から処理を始める。

【0121】(ステップS205-1)信号1「トウキョウス」の音節単位の認識を音声分析部201に依頼する。その結果、モーラ記号列「トオキョオス」が得られたとする。

【0122】(ステップS205-2)信号「トウキョウス」の認識結果として「東京」が得られているので、ステップS205-4へ進む。

【0123】(ステップS205-4)モーラ記号列「トオキョオス」のモーラ記号列長は5である。また、部分認識結果「東京」は置換表現記憶部206に図21のように記録されていたとする。

【0124】このモーラ記号列長とを比較して、入力された信号の部分音節認識結果「トオキョオス」の方が長いので、ステップS205-6へ進む。

【0125】(ステップS205-6)音節認識結果「トオキョオス」と部分認識結果「東京」のモーラ記号列「トオキョオ」を比較すると、図16のようになり、モーラ記号「ス」が余りとして検出される。

【0126】(ステップS205-7)余りとしてモーラ記号「ス」が検出されたので、ステップS205-8へ進む。

【0127】(ステップS205-8)余りのモーラ記号「ス」は音節認識結果「トオキョオス」の最後部に位置し、また、この音節認識結果の元となる信号「トウキョウス」はワイルドカード表現部分「ナントカ」の直前

にあるので、余り「ス」はワイルドカード表現の一部と判断される。ステップS205-9に進む。

【0128】(ステップS205-9)ワイルドカード表現の前部の発音をためるバッファにモーラ記号「ス」を入力する。

【0129】次に、信号[ホテル]について同様の処理を行う。ここでは、部分認識結果「ホテル」の他の余り部分を見つけることができなかつたとし、バッファには何も記録せずに次の処理に進む。

【0130】(ステップS206)ここでは、これまでの情報から適切な語彙を検索する。

【0131】(ステップS206-1)原信号情報と部分認識結果や、ワイルドカード表現の位置情報などから音声入力された対象となる語彙は「東京(ワイルドカード表現)ホテル」と判断される。語彙記憶部202に記録されている語彙から、上記の条件に合う適切な語彙を抽出すると、「東京ステインホテル」、「東京丸の口ホテル」、「東京エンターコンチネンタルホテル」が選択される。また、これらの条件から置換表現記憶部206に登録されている表現からワイルドカード表現で代替された表現として、「ステイン」、「丸の口」、「エンターコンチネンタル」が出力候補として選択される。この時点で、「パルスホテル」が出力候補から出力候補から外されるか、下位の候補となる。

【0132】(ステップS206-2)ステップS205で記録されたバッファを参照すると、モーラ記号「ス」から始まる表現の「ステイン」が有力であると判断できる。ここで、出力候補として、「ステイン」が含まれた語彙「東京ステインホテル」が有力となる。「東京丸の口ホテル」「東京エンターコンチネンタルホテル」は出力候補から外されるか、下位の候補となる。

【0133】(ステップS206-3)ワイルドカード表現検出部203から送られた情報から使用されたワイルドカード表現(「ナントカ」)は数単語置換語であることが分かるので、「東京ステインホテル」を第1位候補として出力する。あるいは、アプリケーション部分が複数候補にも対応している場合は下位の候補として「東京丸の口ホテル」「東京エンターコンチネンタルホテル」を、更に下位の候補として「パルスホテル」を出力する。そして、アプリケーション特有の処理(図中200)がこの出力を受け、適切な処理を行う。

【0134】以上で「トウキョウスナントカホテル」と音声入力された場合の処理を終了する。

【0135】以上の説明によって、本実施形態に係る音声入力分析装置は、利用者が「東京ステインホテル」という名称を明確に記憶していない状態でも、記憶している部分を具体的に、わからない部分をワイルドカード表現を用いて「東京スなんとかホテル」と音声入力することによって、適切な名称に解釈してアプリケーション

部分に情報を出力することが可能であり、また、利用者の知っている細かい情報「東京ス…ホテル」のわからない部分をワイルドカード表現を利用して「東京スなんとかホテル」と入力することにより、おなじく「東京…ホテル」の形式の名称を持つ「東京丸の口ホテル」、「東京エンターコンチネンタルホテル」よりも「東京ステインホテル」のほうが優先され、利用者の入力した音声情報が有効に利用されていることがわかる。

【0136】かくしてこのように構成された本装置によれば、利用者が正確に発声できる単語あるいは文章を記憶しなくとも動作する音声入力解釈装置を構築できる。

【0137】例えば、利用者が発声可能な単語あるいは文章の一部分のみを記憶している場合でも音声の誤認識をおさえ、音声入力をもつシステムの出力を利用者の意図にそったものと導くことのできる音声入力解釈装置を構築できる。

【0138】また、利用者が発声可能な単語あるいは文章の「リズム」のみを記憶している場合でも音声の誤認識をおさえ、音声入力をもつシステムの出力を利用者の意図にそったものと導くことのできる音声入力解釈装置を構築できる。

【0139】尚、各実施形態の作用効果は上述した例に限定されるものではない。例えば、第1の実施形態では置換表現照合部103、第2の実施形態では置換表現照合部205において置換処理された結果のリストを利用者に提示し、正しいものを選択させることによって誤動作を避けることができる。

【0140】また、マルチモーダルインターフェースの入力手段として利用し、検索幅を更に狭め、出力の冗長をおさえ、利用者の負担を軽減することも可能である。

【0141】また、マルチモーダルインターフェースのみに限らず、任意の音声入力に伴う装置の入力手段として利用することが可能である。また、韻律情報はワイルドカード表現された部分のみに限らず、入力された音声情報すべてに対して解析、利用することも可能である。

【0142】以下では、本音声入力解釈装置における処理をソフトウェアを使って実現する場合の装置構成について図22を参照しながら説明する。

【0143】この場合、本音声入力解釈装置のハードウェア部分は、CPU21、プログラムや必要なデータを格納するためのRAM22、ディスクドライブ装置24、記憶装置25、入出力装置26である。

【0144】第1の実施形態の場合、図1の音声分析部101、語彙記憶部102、置換表現照合部103、韻律情報記憶部104は、それぞれの処理手順を記述したプログラムにより構成される。

【0145】第2の実施形態の場合、図11の音声分析部201、語彙記憶部202、ワイルドカード表現検出部203、ワイルドカード表現記憶部204、置換表現照合部205、置換表現記憶部206は、それぞれの処



理手順を記述したプログラムにより構成される。

【0146】なお、各記憶部に格納する情報は、プログラムと一体化されたものであってもよいし、プログラムとは別に設定されるものであってもよい。

【0147】この処理手順を記述したプログラムは、図22のコンピュータシステムを制御するためのプログラムとしてRAM22に格納され、CPU21により実行させる。CPU21はRAM22に格納されたプログラムの手順に従い、演算や、記憶装置25あるいは入出力装置26の制御などを行って、所望の機能を実現していく。

【0148】プログラムをRAM22にインストールするには種々の方法を用いることができる。例えば、上記プログラム(図1の音声分析部101、語彙記憶部102、置換表現照合部103、韻律情報記憶部104の処理手順を記述したプログラムであって、コンピュータシステムを制御するためのプログラムや、図11の音声分析部201、語彙記憶部202、ワイルドカード表現検出部203、ワイルドカード表現記憶部204、置換表現照合部205、置換表現記憶部206の処理手順を記述したプログラムであって、コンピュータシステムを制御するためのプログラム)を、コンピュータで読みとり可能な記憶媒体(例えばフロッピーディスク、あるいはCD-ROM等のリムーバブル記憶媒体)に記憶させておく。そして、図22に示すように記憶媒体に応じたディスクドライブ装置24を用いて該プログラムを読みとり、RAM22に格納する。あるいは、いったんディスクドライブ装置24等にインストールしておき、実行時に同装置からRAM22に格納する。

【0149】また、プログラムを格納した記憶媒体がICカードである場合は、ICカードリーダーを用いて該プログラムを読みとることができる。さらには、ネットワークを介して所定のインターフェース装置からプログラムを受けとることもできる。

【0150】なお、音声入力解釈装置にその解釈結果を利用するアプリケーションを搭載してもよいし、音声入力解釈装置とアプリケーションを搭載する装置を独立したものにしてもよい。また、音声入力解釈装置を実現するプログラムとその解釈結果を利用するアプリケーションを実現するプログラムとを、同一のCPU上で実行してもよいし、別々に設けたCPU上で実行してもよい。

【0151】ところで、第1、第2の実施形態では、ワイルドカード表現が1つしか入力されないという前提で実現しているように記述しているが、ワイルドカード表現が複数の入力が行われても、第1の実施形態では対応する語彙を語彙記憶部102に生成し、置換表現照合部103においては該当するワイルドカード表現部分のそれぞれについて同様の処理を行えば扱うことが可能であり、また第2の実施形態では複数検出されたワイルドカード表現について、その位置と、種類、韻律に関する情

報を置換表現照合部205に渡し、また、ワイルドカード表現の一部を記録するバッファをワイルドカード表現の途中を記録するためのものを追加し、連続してワイルドカード表現が現れた場合はまとめて1つのワイルドカード表現として、検出された各ワイルドカード表現について同様に処理を行えば扱うことが可能である。

【0152】また、第1、第2の実施形態で設定される検索条件は特に各実施形態に固有のものではなく、例えば、第2の実施形態における置換表現検索時に音声入力時間を利用しても良い。また、第1の実施形態については「『ス』なんとか」のようにワイルドカード表現の一部に正しい表現を交えた入力はされないという前提で実現しているように記述しているが、語彙記憶部102に「す/なんとか」のような語彙を設定すれば容易に対応可能である。また、ワイルドカード表現を数単語置換語や、リズム語に定義しなくとも、全ての表現について韻律などを検索条件にすることも可能である。

【0153】また、日本語に限らず、ワイルドカード表現が存在する言語全てにモーラ記号単位の分析を音節あるいは音素などの共通の単位の分析とすることによって、本発明を適用することが可能である。また、本発明を例えば歌詞の分からない部分をリズムで歌う入力によって音楽の検索に適用することも可能である。

【0154】本発明は、上述した実施の形態に限定されるものではなく、その技術的範囲において種々変形して実施することができる。

【0155】

【発明の効果】本発明によれば、入力音声から正規の語彙の一部を代替表現した部分を検出しこの部分に妥当する正規の表現に置換するので、音声入力として許容される語彙を利用者が明確に覚えなくとも、その代替表現を含む音声入力を受け入れ、これを解釈することができる。

【図面の簡単な説明】

【図1】本発明の第1の実施形態に係る音声入力解釈装置の構成例を示す図

【図2】語彙記憶部に記録される情報の一例を示す図

【図3】語彙記憶部に記録される語彙の一例を示す図

【図4】音声分析部から置換表現照合部へ渡される情報の一例を示す図

【図5】韻律情報記憶部に記録されている情報の一例を示す図

【図6】置換表現照合部の動作の一例を示すフローチャート

【図7】語彙記憶部に登録された語彙の一例を示す図

【図8】韻律情報記憶部に登録された情報の一例を示す図

【図9】音声分析部から置換表現照合部へ出力する情報の一例を示す図

【図10】音声認識結果に適合する語彙の検索結果の一

【図22】ハードウェア構成の一例を示す図

【符号の説明】

1, 2…音声入力解釈装置

100…音声入力装置

101…音声分析部

102…語彙記憶部

103…置換表現照合部

104…韻律情報記憶部

201...音声分析部

202...語彙記憶部

203…ワイルドカード表現検出部

204…ワイルドカード表現記憶部

205…置換表現照合部

206...置換表現記憶部

21...CPU

22...RAM

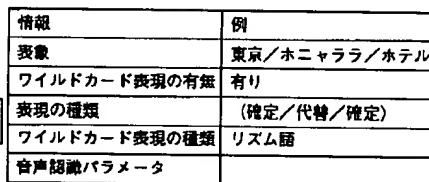
23...バス

## 24…ディスクドライブ装置


## 25…記憶装置

26…入出力装置

【図2】




【図4】





| 情報           | 例  |
|--------------|--|
| 認識結果         | 「東京／ホニャララ／ホテル」   |
| 単語発声時間       | (660msec/820msec/510msec)  |
| 韻律パラメータ      |  |
| ワイルドカード表現の有無 | 有り   |
| 表現の種類        | (確定／代替／確定)   |
| ワイルドカード表現の種類 | リズム図   |

【図3】

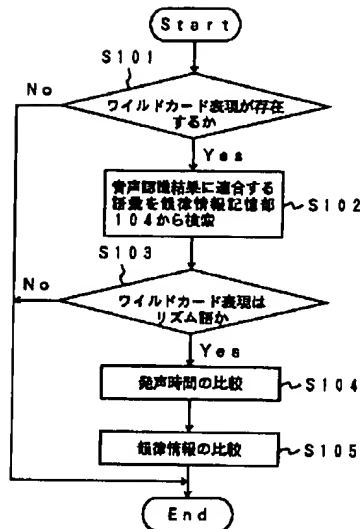
|               |               |
|---------------|---------------|
| 東京/ステイン/ホテル   | なんとか/ステイン/ホテル |
| 東京/なんとか/ホテル   | 東京/ステイン/なんとか  |
| なんとか/ホテル      | 東京/なんとか       |
| ホニャラ/ステイン/ホテル | 東京/ホニャララ/ホテル  |
| 東京/ステイン/ホニャラ  | ホニャララララララ/ホテル |
| 東京/ホニャラララララ   |               |

【図8】

|      |   |
|------|---|
| 情報   | 例   |
| 表像   | 「東京/ステイン/ホテル」   |
| 標準時間 | (700msec/900msec/550msec)   |
| 韻律   |  |

|      |   |   |
|------|---|---|
| 語彙   | 「パルスホテル」  | 「東京ステインホテル」   |
| 表象   | パルス／ホテル   | 東京／ステイン／ホテル   |
| 標準時間 | (450msec/550msec)   | (700msec/900msec/550msec)   |
| 韻律情報 |  |  |
| 語彙   | 「東京丸の口ホテル」  | 「東京エンターコンチネンタルホテル」  |
| 表象   | 東京／丸の口／ホテル  | 東京／エンター／コンチネンタル／ホテル   |
| 標準時間 | (690msec/850msec/560msec)   | (700msec/650msec/1050msec/560msec)  |
| 韻律情報 |  |  |

【図6】



【図7】

|                      |                       |
|----------------------|-----------------------|
| パルス/ホテル              | 東京/ステイン/ホテル           |
| 東京/丸の口/ホテル           | 東京/エンター/コンチネンタル/ホテル   |
| ホニャラ/ホテル             | パルス/ホニャラ              |
| ホニャラ/ステイン/ホテル        | 東京/ホニャララ/ホテル          |
| 東京/ステイン/ホニャラ         | ホニャラ/丸の口/ホテル          |
| 東京/丸の口/ホニャラ          | ホニャラララララ/ホテル          |
| 東京/ホニャラララララ          | ホニャラ/エンター/コンチネンタル/ホテル |
| 東京/ホニャラ/コンチネンタル/ホテル  | 東京/エンター/ホニャララララ/ホテル   |
| 東京/エンター/コンチネンタル/ホニャラ | ホニャラララララ/コンチネンタル/ホテル  |
| 東京/ホニャラララララララ/ホテル    | 東京/エンター/ホニャラララララララ    |
| ホニャララララララララララ/ホテル    | 東京/ホニャララララララララララ      |

【図10】

|                  | 比率   |       | 伸長係数   | 伸長後<br>(msec) | 標準時間<br>(msec) |
|------------------|------|-------|--------|---------------|----------------|
|                  | 「東京」 | 「ホテル」 |        |               |                |
| 東京ステインホテル        | 700  | 550   | 1.0777 | 884           | 900            |
|                  | 650  | 510   |        |               |                |
| 東京丸の口ホテル         | 690  | 580   | 1.0798 | 885           | 850            |
|                  | 650  | 510   |        |               |                |
| 東京エンターコンチネンタルホテル | 700  | 560   | 1.0875 | 892           | 1700           |
|                  | 650  | 510   |        |               |                |

【図12】

| 情報            | 例            |
|---------------|--------------|
| 表文字列          | 東京ステインホテル    |
| モーラ記号列        | トウキョオステインホテル |
| モーラ記号長        | 12           |
| 韻律パラメータ       | →→→→→→→→→→→→ |
| 音声認識に必要なパラメータ |              |

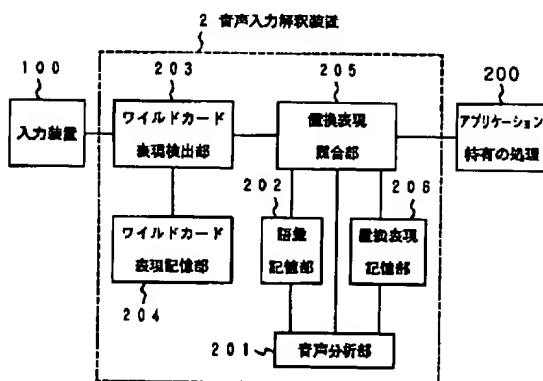
【図13】

| 情報                                     | 例                  |
|--|--------------------|
| Wildcard expression existence          | 有り                 |
| 原信号                                    | 「トウキョウ」「ホニャラ」「ホテル」 |
| Wildcard expression position           | 2                  |
| Wildcard expression type               | リズム語               |
| Wildcard expression mora code length   | 4                  |
| Wildcard expression rhythm information | ↗→→→               |

【図9】

|                               |                           |
|-------------------------------|---------------------------|
| 認識結果                          | 「東京/ホニャララ/ホテル」            |
| 単語発声時間                        | (650msec/820msec/510msec) |
| 韻律パラメータ                       | →→→→→→→→→→→→              |
| Wildcard expression existence | 有り                        |
| 表現の種類                         | (確定/代替/確定)                |
| Wildcard expression type      | リズム語                      |

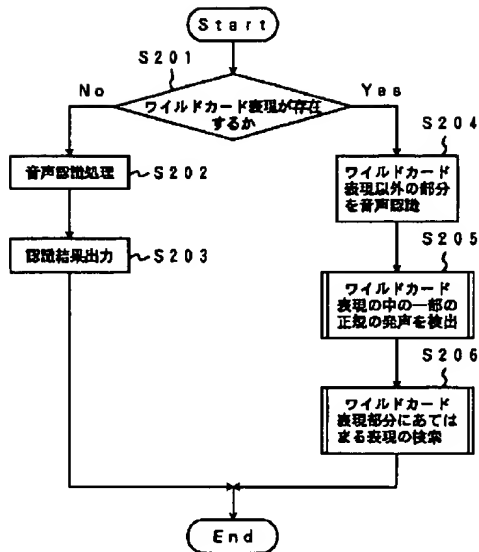
【図11】



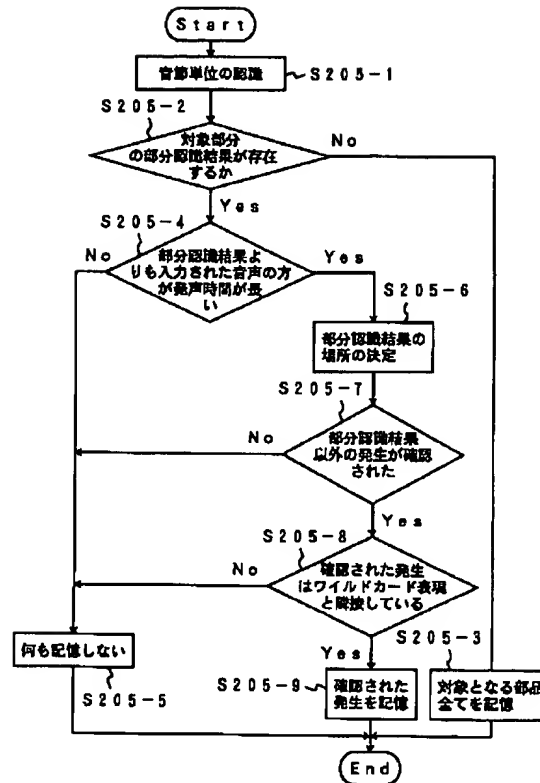
【図21】

|         |       |
|---------|-------|
| 表文字     | 東京    |
| モーラ記号列  | トウキョオ |
| モーラ記号列長 | 4     |
| 韻律パラメータ | →→→→  |

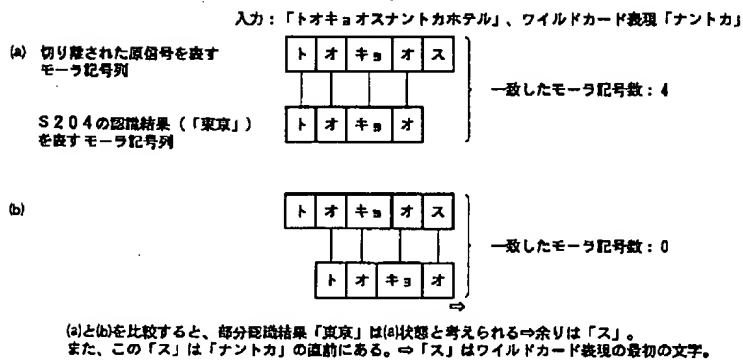
【図14】



【図15】



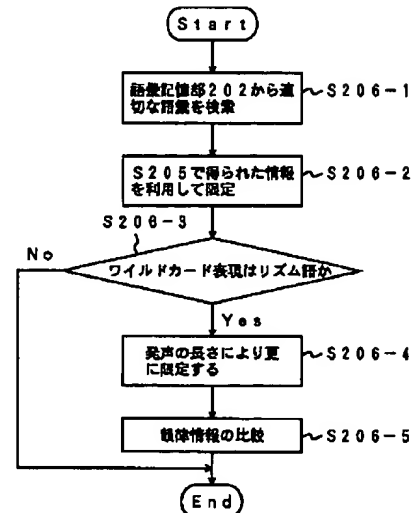
【図16】



【図19】

|            |               |                |
|------------|---------------|----------------|
| 東京         | ステイン          | ホテル            |
| 東京ステイン     | ステインホテル       | 丸の口            |
| 東京丸の口      | 丸の口ホテル        | エンター           |
| コンチネンタル    | 東京エンター        | エンターコンチネンタル    |
| コンチネンタルホテル | 東京エンターコンチネンタル | エンターコンチネンタルホテル |
| パルス        |               |                |

【図17】



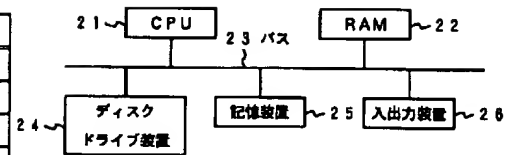
【図18】

|         | 東京ステインホテル              | 東京丸の口ホテル         |
|---------|------------------------|------------------|
| 表像文字列   | 東京ステインホテル              | 東京丸の口ホテル         |
| モーラ記号列  | トオキヨステインホテル            | トオキヨマルノクチホテル     |
| モーラ記号長  | 12                     | 12               |
| 韻律パラメータ | →→→→→→→→→→→→→→→→       | →→→→→→→→→→→→→→→→ |
|         | 東京エンターコンチネンタルホテル       | パルスホテル           |
| 表像文字列   | 東京エンターコンチネンタルホテル       | パルスホテル           |
| モーラ記号列  | トオキヨエンタコンチネンタルホテル      | パルスホテル           |
| モーラ記号長  | 18                     | 6                |
| 韻律パラメータ | →→→→→→→→→→→→→→→→→→→→→→ | →→→→→→→→→→→→→→→→ |

【図20】

| 情報の種類          | 内 容                   |
|----------------|-----------------------|
| ワイルドカード表現の有無   | 有り                    |
| 原番号            | 【トウキョウス】 【ナントカ】 【ホテル】 |
| ワイルドカード表現の位置   | 2                     |
| ワイルドカード表現の種類   | 数単語置換語                |
| ワイルドカード表現の文字列長 | —                     |
| ワイルドカード表現の韻律情報 | —                     |

【図22】



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**